



Géolocalisation et prédiction dans les réseaux Wi-Fi en intérieur

Frédéric Lassabe

► To cite this version:

Frédéric Lassabe. Géolocalisation et prédiction dans les réseaux Wi-Fi en intérieur. Réseaux et télécommunications [cs.NI]. Université de Franche-Comté, 2009. Français. NNT : . tel-00396265

HAL Id: tel-00396265

<https://theses.hal.science/tel-00396265>

Submitted on 17 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Géolocalisation et prédiction dans les réseaux Wi-Fi en intérieur

MÉMOIRE

soutenu le 21 Avril 2009

pour l'obtention du grade de docteur

Thèse de l'Université de Franche-Comté–Besançon
(Informatique, Automatique et Productique)

par

Frédéric LASSABE

Composition du jury

Francois Spies, professeur à l'Université de Franche-Comté
Philippe Canalda, maître de conférence à l'Université de Franche-Comté
Pascal Chatonnay, maître de conférence à l'Université de Franche-Comté
David Simplot-Ryl, professeur à l'Université de Lille 1
Thomas Noël, professeur à l'Université de Strasbourg
Michel Banâtre, directeur de recherche INRIA

Remerciements

Le bon déroulement de ma thèse et la rédaction de ce mémoire n'auraient pu avoir lieu sans le concours de nombreuses personnes que je tiens à remercier :

François SPIES, Professeur au LIFC, qui m'a accueilli au sein de l'équipe de recherche de Montbéliard et suivi tout au long de la thèse.

Pascal CHATONNAY et Philippe CANALDA, Maîtres de conférences au LIFC qui ont su m'aider et qui m'ont apporté leur soutien tout au long de la thèse.

Julien BOURGEOIS, Gérard CÉCÉ, Damien CHARLET et Eugen DEDU, membres de l'IUT Belfort-Montbéliard pour leur sympathie ainsi que pour leur accueil sur le site de Montbéliard.

David SIMPLOT-RYL, Professeur au LIFL, pour avoir accepté de présider le jury de la thèse.

Michel BANÂTRE, directeur de recherche à l'INRIA Rennes, et Thomas NOËL, Professeur à l'Université de Strasbourg, pour avoir accepté le rôle de rapporteurs et de faire partie du jury de la thèse.

L'ensemble des doctorants du LIFC en particulier Jean-Baptiste ERNST-DESMULIER pour leur patience et leur aide.

Un merci tout particulier à Sébastien LINCK, également doctorant au LIFC, pour m'avoir supporté pendant ma thèse.

Table des matières

Introduction	1
I Géolocalisation de terminaux mobiles dans les réseaux Wi-Fi	5
1 État de l’art de la géolocalisation dans les réseaux Wi-Fi	7
1.1 Terminologie de l’infrastructure des systèmes de géolocalisation	8
1.2 Classification des systèmes de géolocalisation	9
1.2.1 Critères liés aux techniques et méthodes de géolocalisation	9
1.2.2 Symbologie	10
1.3 Géolocalisation discrète	11
1.3.1 Techniques de géolocalisation cellulaire	11
1.3.2 Communications courte portée	12
1.3.3 Usage des ultrasons	12
1.4 Géolocalisation continue	13
1.4.1 Positionnement par satellite	13
1.4.2 Utilisation de capteurs	14
1.5 Géolocalisation des terminaux Wi-Fi	15
1.5.1 Collecte de mesures des caractéristiques des signaux reçus	15
1.5.2 Utilisation d’un modèle de propagation	21
1.5.3 Réseaux de Neurones	24
1.6 Analyse	25
2 Système de géolocalisation dans les réseaux Wi-Fi en intérieur	29
2.1 Caractéristiques de l’intérieur des bâtiments	30
2.1.1 Observations pratiques : le cas de Numérica	31
2.1.2 Topologie hétérogène et découpage en zones homogènes	34
2.2 Calibration du modèle de propagation des ondes radio	35
2.2.1 Modèle de Friis calibré	36

2.2.2	De l'analyse de l'atténuation du signal à la géolocalisation	37
2.2.3	Limites	40
2.3	Prise en compte de la topologie dans les algorithmes de type Viterbi	40
2.3.1	Représentation de la topologie	41
2.3.2	Réduction de la complexité de l'algorithme Viterbi-like	43
2.4	Fusion de la cartographie des puissances et du FBCM	45
2.4.1	Mesures de référence	45
2.4.2	Fonctionnement du FRBHM	45
2.4.3	Algorithmes FRBHM	46
2.4.4	Calcul de la matrice des distances du FVLI dans le FRBHM continu	49
2.5	Expérimentation et comparaison	50
2.5.1	Contextes des expérimentations	51
2.5.2	Protocole de test	53
2.5.3	Résultats et analyse	53

II Prédiction de la mobilité 59

3	Modèles d'apprentissage des transitions	61
3.1	Définition des horizons temporels	62
3.1.1	Modalité de déplacement	63
3.1.2	Court terme	64
3.1.3	Moyen terme	64
3.1.4	Long terme	64
3.2	Prédiction des trajectoires	64
3.3	Modèles d'apprentissage états-transitions	66
3.3.1	Modèles de Markov	67
3.3.2	Découverte des voisins	71
3.3.3	Correction des erreurs d'observation	72
3.3.4	Inférence Bayésienne	73
3.3.5	Autres modèles	75
4	Prédiction de la mobilité dans les réseaux multimédia sans-fil	77
4.1	Problématique et choix	78
4.2	Journalisation et conservation des données	80
4.2.1	Choix des données	80
4.2.2	Problèmes potentiels et leurs solutions	81
4.3	Seuils de prédiction	81

4.3.1	Contexte de la continuité des services	81
4.3.2	Définition de seuils de préfetching	82
4.4	Modèles proposés	82
4.4.1	Modèles de Markov et mobilité, sans prise en compte de la temporalité	82
4.4.2	Modèle k-past : historique de longueur constante	83
4.4.3	k-to-1-past, k-to-1-past* : historiques de tailles variables	86
4.4.4	Prise en compte de la date et des délais	87
4.5	Validation des modèles de prédiction de la mobilité	88
4.5.1	Données de cas réels et simulation	88
4.5.2	Expérimentations	89
III	Mise en œuvre	97
5	Mise en œuvre du système de géolocalisation	99
5.1	GeoMoVie	99
5.1.1	Contexte	100
5.1.2	NetMoVie	100
5.1.3	SysMoVie	100
5.1.4	GeoMoVie	101
5.2	GuiNuMo	101
5.3	Architecture	103
5.3.1	Choix d'une approche orientée service	103
5.3.2	Architecture matérielle	103
5.3.3	Architecture logicielle	104
	Conclusion	109
IV	Annexes	113
A	Notions fondamentales	115
A.1	Ondes radio	115
A.1.1	Unités de mesure du signal	115
A.1.2	Phénomènes de trajectoire	117
A.1.3	Atténuation : la formule de Friis	119
A.2	Calcul de positionnement	120
A.2.1	Triangulation	120
A.2.2	Trilatération	121

A.3	La norme 802.11	122
A.3.1	Accès au médium de transmission	122
A.3.2	Les révisions de la norme 802.11	124
A.4	Modèles de Markov	125
A.4.1	Chaînes de Markov à espace d'états discret	125
A.4.2	Chaînes de Markov d'ordres supérieurs à 1	125
A.4.3	Modèle de Markov à espace d'états caché	125
A.4.4	Algorithme de Viterbi	127
A.5	Inférence Bayésienne	128

Bibliographie	131
----------------------	------------

Liste des tableaux

2.1	Calcul d'indices de Friis et de distances, AP UFC-DEA.	37
2.2	Comparaison des complexités des algorithmes Viterbi-like et Viterbi-like-improved.	44
2.3	Résultat des expérimentations, mesures effectuées par le terminal mobile.	54
2.4	Résultat des expérimentations, mesures effectuées par l'infrastructure en dispo- sant de 3 AP.	54
2.5	Résultat des expérimentations, mesures effectuées par l'infrastructure en dispo- sant de 4 AP.	54
2.6	Résultat des expérimentations, mesures effectuées par l'infrastructure en dispo- sant de 5 AP.	55
2.7	Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l'infrastructure en disposant de 3 AP.	55
2.8	Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l'infrastructure en disposant de 4 AP.	55
2.9	Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l'infrastructure en disposant de 5 AP.	56
3.1	Quantification des propriétés des modes de déplacement.	63
3.2	Exemple de séquences d'états physiques pour l'apprentissage.	67
3.3	Modèle bayésien, résultat complet, sans apprentissage préalable.	75
3.4	Modèle bayésien, prise en compte du temps.	75
A.1	Table des probabilités conditionnelles.	128

Table des figures

1.1	Illustration des définitions et rôles dans la géolocalisation.	8
1.2	Algorithme Viterbi-like dans le système RADAR.	16
1.3	Résultats du système RADAR (erreurs en mètres).	17
1.4	Résultats d'Ekahau (erreurs en mètres).	19
1.5	Tableau récapitulatif des systèmes de géolocalisation Wi-Fi les plus importants. .	26
2.1	Plan 2D-isométrique des locaux du laboratoire LIFC, Numerica rez-de-chaussée et 1 ^{er} étage.	31
2.2	Atténuation du signal dans un couloir, puissance en fonction de la distance. . . .	32
2.3	Comparaison du calcul de la distance : Friis et Interlink Networks.	33
2.4	Disparité des trajectoires au sein des zones homogènes.	35
2.5	Positionnement des points de calibration et des points d'accès du banc d'essai n°1.	36
2.6	Trilatération avec des distances évaluées et biaisées.	38
2.7	Algorithme de calcul du point le plus proche aux cercles.	39
2.8	Représentation du calcul des distances en fonction de la topologie.	41
2.9	Calcul des distances topologiques minimales entre les points de référence.	42
2.10	Algorithme FRBHM basique.	47
2.11	Algorithme FRBHM avec prise en compte de la topologie.	47
2.12	Algorithme FRBHM continu.	49
2.13	Algorithme du calcul des distances entre points quelconques.	50
2.14	Disposition des points d'accès lors des expérimentations.	52
3.1	Horizons temporels de prédiction de la mobilité.	63
3.2	Modèle de Markov du premier ordre.	68
3.3	Modèles de Markov du second et troisième ordres.	69
3.4	Results for LRS with various sizes.	70
3.5	Utilisation des HMM dans la prédiction de la mobilité.	73
3.6	Graphe des dépendances du réseau bayésien de prédiction de la mobilité.	74
4.1	Exemple de données de journalisation.	83
4.2	Construction du modèle de Markov d'ordre K.	84
4.3	Prédiction de la mobilité, modèle k-past.	85
4.4	Prédiction de la mobilité, modèle k-to-1-past.	86
4.5	Prédiction de la mobilité, modèle k-to-1-past*.	87

4.6	Modèle k-past, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.	90
4.7	Modèle k-to-1-past, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.	91
4.8	Modèle k-to-1-past*, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.	92
4.9	Tests des modèles atemporaux : apprentissage sur le profil fall, test sur le profil summer.	93
4.10	Précision avec prise en compte du délai (75/25).	94
5.1	Architecture matérielle du service de gestion de la mobilité.	104
5.2	Architecture fonctionnelle de la plate-forme de test : classes utilisées.	105
5.3	Architecture logicielle du service de gestion de la mobilité : algorithmes.	106
5.4	Mesure effectuée par le terminal mobile.	107
5.5	Mesure effectuée par l'infrastructure.	108
A.1	Diagrammes de couverture d'antennes.	116
A.2	Réflexion d'une onde sur les parois d'un bâtiment.	117
A.3	Réfraction d'une onde pénétrant dans un milieu différent.	118
A.4	Phénomène de diffraction.	119
A.5	Atténuation du signal suivant la formule de Friis.	120
A.6	Principe de la triangulation	121
A.7	Principe de la trilatération.	122
A.8	Configuration des nœuds cachés	123
A.9	Algorithme de Viterbi.	127
A.10	Graphe de dépendance du réseau bayésien "risque d'accident".	128

Glossaire

3G : Troisième génération de téléphonie mobile, permet des débits élevés et l'accès à des services Internet et multimédia.

AKMM : All-Kth Markov Model, combinaison des modèles de Markov d'ordres 1 à K.

AP : Access Point, c'est-à-dire le point d'accès entre le réseau câblé et le réseau sans fil.

BTS : Base Transceiver Station, antenne par laquelle le terminal se connecte au réseau dans les réseaux GSM.

dBi : Décibel isotrope, permet de mesurer le gain d'antenne par rapport à une antenne isotrope, c'est-à-dire rayonnant uniformément.

dBm : Décibel ramené au milliwatt, unité de mesure de la puissance du signal.

ESSID : Extended Service Set Identifier, définit un ou plusieurs points d'accès et leur zone de couverture.

FBCM : Friis-Based Calibrated Model, modèle de Friis calibré.

FRBHM : FBCM and Reference-Based Hybrid Model, modèle de localisation qui fusionne les approches de cartographie des puissances et de propagation du signal.

FVLI : Fast Viterbi-Like Improved, algorithme Viterbi-like optimisé en complexité, prenant en compte la topologie.

GNSS : Global Navigation Satellite System, système satellite global de navigation. Système permettant la localisation sur la Terre à l'aide de satellites.

GPS : Global Positioning System. GNSS développé par le département de la Défense des U.S.A.

GSM : Global System for Mobile Communications, norme numérique de seconde génération pour la téléphonie mobile.

handoff : ou Handover, changement d'équipement de connexion au réseau.

HMM : Hidden Markov Model, modèle de Markov dont les observations correspondent à des états suivant une loi de probabilité.

IN : Interlink Networks, propose une variante de Friis globale pour tous les bâtiments.

KMM : Modèle de Markov d'ordre K, étiquetant un état par K observations.

LIFC : Laboratoire d'Informatique de l'Université de Franche-Comté, où cette thèse a été réalisée.

MLP : Multi Layer Perceptron, un type de réseau de neurones.

MM : Modèle de Markov, processus stochastique sans mémoire.

MPEG-2 : Norme d'encodage et décodage vidéo du Moving Picture Experts Group, utilisée notamment dans les DVD.

PDA : Personal Digital Assistant, ou assistant personnel numérique.

PIRE : Puissance Isotrope Rayonnée Équivalente, indique la puissance émise par l'antenne relativement à une antenne isotrope.

RFID : Radio Frequency IDentification, puce qui contient des informations et les transmet par ondes radio. Il existe de nombreuses variétés de puces RFID.

RTAP : Radio Tap, un outil de mesure des signaux radio intégré au pilote de certaines cartes Wi-Fi.

SNAP-WPS : Satellite Navigation and Positioning Group - Wireless Positioning System propose le calcul de distance en fonction de la puissance en se basant sur une régression polynomiale par rapport à des mesures de calibration.

SS : Signal Strength, la mesure de la puissance du signal.

VOD : Video On Demand, service de diffusion de vidéo à l'utilisateur dans lequel celui-ci sélectionne la vidéo de son choix.

Wi-Fi : Wireless Fidelity, une implantation de la norme IEEE802.11 basée sur des transmissions radio.

WLAN : Wireless Local Area Network, réseau local sans-fil.

Introduction

L'accroissement de la connectivité des terminaux informatiques fait apparaître de nouvelles perspectives d'applications. Auparavant, les terminaux utilisaient des supports locaux, tels que le CD et le DVD, pour consulter des contenus multimédia de qualité. Un DVD encodé en MPEG-2 requiert un débit de 15 Mbps. L'augmentation des débits permet d'atteindre 1 Gbps en réseaux locaux câblés, 54 Mbps en communications sans-fil et 25 Mbps sur les connexions Internet grand-public avec l'ADSL. De tels débits rendent les contenus multimédia consultables en ligne, dans des conditions proches ou égales à celles DVD. Une application rendue possible par l'augmentation des débits disponibles est la vidéo à la demande (*Video On Demand*, VOD). Grâce à ce service, un utilisateur se connecte à un réseau pour choisir la vidéo qu'il souhaite visionner.

La plus grande évolution des débits des réseaux est visible dans les réseaux sans-fil. Alors que le GSM offre un débit de 9,6 kbps juste suffisant pour la voix, la téléphonie mobile a évolué vers les réseaux 3G, dont les débits sont suffisants pour transmettre des contenus multimédia. Parallèlement, les réseaux informatiques se sont dotés d'appareils sans-fil. La norme la plus utilisée est actuellement la norme IEEE 802.11. Elle permet, dans sa révision 802.11g, des débits de 54 Mbps.

Ces deux évolutions majeures des matériels réseau et terminaux mènent à de nouvelles formes d'informatique. De l'informatique centrée sur la machine, généralement figée à un emplacement précis, nous nous orientons aujourd'hui vers une nouvelle conception : l'informatique diffuse. L'informatique diffuse est centrée sur les usagers et les contenus plutôt que sur le matériel qui permet la fourniture du service à l'utilisateur. Dans l'informatique diffuse, l'information est disponible partout, avec n'importe quel appareil communicant (ordinateur, téléphone mobile, PDA). De plus, l'information disponible est riche car composée de média variés tels que de la vidéo, du son, du texte. Enfin, les contenus fournis par les systèmes informatiques diffus sont adaptés à la situation de l'utilisateur du service [1]. L'adaptation est la dépendance au contexte de l'utilisateur et on parle donc d'applications *context-aware*.

Problématiques

Au cours des travaux, deux problématiques majeures ont été déterminées. Il s'agit, d'une part, de la géolocalisation des terminaux mobiles dans un réseau Wi-Fi et, d'autre part, la prédiction de la mobilité des terminaux.

Positionnement de terminaux mobiles dans un réseau Wi-Fi en intérieur

Pour permettre ces caractéristiques des services, deux problématiques majeures doivent être résolues. La première est l'acquisition du contexte de l'utilisateur, détecté par son terminal mobile. En particulier, une donnée fondamentale du contexte de l'utilisateur est sa position géographique. Déterminer sa position est critique pour le fonctionnement d'une application dépendante du contexte. On nomme cette opération la "géolocalisation". En extérieur, la géolocalisation est rendue possible par les systèmes satellitaires (*Global Navigation Satellite System* - GNSS). Cependant, en intérieur ainsi que dans des environnements couverts (forêt et urbanisme dense par exemple), les systèmes satellitaires deviennent inopérants. Il est nécessaire de leur trouver un système de substitution.

Continuité de services en mobilité

La seconde problématique est la continuité des services fournis à l'utilisateur, via son terminal mobile. En effet, la mobilité des terminaux dans les réseaux informatiques diffus conduit à la modification de la topologie du réseau quand un terminal change de point d'accès au réseau d'infrastructure. Lorsqu'un terminal modifie ainsi la topologie du réseau, il subit non seulement des déconnexions le temps de basculer d'un point d'accès au suivant, mais peut également changer de sous-réseau si ses deux points d'accès sont situés dans des sous-réseaux différents. Dans ce dernier cas, il est également nécessaire que les fournisseurs du service changent l'adresse du terminal auquel les données sont acheminées et que le routage sur le réseau soit mis à jour. Toutes ces actions coûtent du temps pendant lequel le terminal mobile est déconnecté de ses services et l'utilisateur ne reçoit plus ce qu'il est en droit d'attendre comme services.

Nos travaux prennent place dans le contexte de l'informatique diffuse sur des réseaux Wi-Fi. En effet, ceux-ci sont largement développés et présents, que ce soit en entreprise ou chez les particuliers. De plus, les débits proposés par les différentes révisions de la norme 802.11, sous-jacente aux réseaux Wi-Fi, permettent d'envisager la diffusion de contenus multimédia riches. Nous considérons des services qui fournissent ce type de contenus riches à des terminaux mobiles, susceptibles de changer de point d'accès au réseau au cours de leurs pérégrinations. La gestion de la mobilité et la géolocalisation doivent être transparentes pour l'utilisateur et s'adapter aux conditions. En particulier, à moyen terme, il est souhaitable que le système de géolocalisation se mette en œuvre avec le minimum d'interventions humaines. Il est également souhaitable que les coordonnées obtenues puissent être mises en relation avec des coordonnées géocentriques absolues, de façon à permettre la continuité de la géolocalisation entre le système Wi-Fi et les GNSS.

Thèse défendue

Dans ce document, je soutiens la thèse que la gestion de la mobilité dans un réseau Wi-Fi s'appuie sur la géolocalisation et la gestion pro-active des procédures de handoff, basée sur la prédiction de la mobilité. La géolocalisation est effectuée par un système hybride, combinant les approches discrètes et continues. La gestion pro-active du *handoff* est rendue possible par l'apprentissage des pérégrinations des terminaux mobiles et leur représentation dans un modèle statistique. Le *handoff* est l'ensemble des opérations nécessaires au basculement du terminal

mobile d'un point d'accès au réseau à un autre point d'accès sans interrompre ses services. Le modèle est exploité pour prédire où un terminal mobile va se diriger prochainement et ainsi préparer sa procédure de handoff.

Plan du mémoire

Deux parties regroupent le cœur des travaux. La première partie traite de la géolocalisation des terminaux mobiles dans un réseau Wi-Fi. En effet, nous nous plaçons dans le contexte d'un réseau Wi-Fi, dans lequel les terminaux sont mobiles. Le contexte de nos travaux étant la localisation à l'intérieur des bâtiments, nous ne pouvons recourir aux systèmes de localisation par satellite. C'est pourquoi nous proposons notre propre système de géolocalisation.

Le premier chapitre de la partie traitant la géolocalisation expose les travaux dans le domaine de la géolocalisation, en particulier lorsqu'elle concerne les réseaux Wi-Fi déployés en intérieur. Ces travaux nous ont permis de classer les systèmes selon leur fonctionnement et d'en déduire les particularités exploitables dans notre contexte. En particulier, nous nous sommes intéressés à la prise en compte de la topologie par le système de géolocalisation. Nous avons identifié deux grands axes de travail concernant la géolocalisation des terminaux Wi-Fi en intérieur : le calcul de la position à l'aide d'un modèle de propagation des ondes radio et l'utilisation d'une cartographie des puissances.

Le chapitre suivant décrit nos contributions dans le domaine de la géolocalisation de terminaux mobiles dans les réseaux Wi-Fi en intérieur. Nous y présentons une méthode de calibration d'un modèle de propagation des ondes afin d'obtenir un modèle adapté à son lieu de déploiement. Ce modèle est exploité dans un système de géolocalisation basé sur un découpage du territoire en zones de propriétés homogènes. Une méthode de cartographie des puissances permet de déterminer quel modèle sera appliqué en fonction de la détection du terminal mobile dans une zone homogène. Enfin, nous proposons d'affiner le résultat de la géolocalisation en éliminant des positions ambiguës grâce à l'historique des positions d'un terminal mobile et à un plan de déplacement des terminaux dans le bâtiment.

La seconde partie traite de la prédiction de la mobilité. Dans un premier chapitre, nous présentons des modèles permettant de rendre compte d'un changement d'état dans un système. Nous écartons les calculs de trajectoires, trop imprécis pour les terminaux mobiles piétons, au profit des modèles discrets. En particulier, les approches par modèles de Markov, largement exploités dans le domaine de la prédiction de chargement de documents, et les réseaux bayésiens sont intéressants pour nous. Nous nous intéressons à des modèles discrets car ils permettent de rendre compte du passage d'un état à un autre. Ceci correspond dans notre cas à l'étude des changements de zone de couverture. Les zones de couvertures concernent soit la couverture réseau, par exemple par les points d'accès dans un réseau Wi-Fi, soit la couverture des services, par exemple la zone desservie par un cache de données.

Le second chapitre présente nos contributions. Nous y décrivons dans un premier temps des modèles atemporaux, basés sur des combinaisons de degrés de modèles de Markov. Dans un second temps, pour répondre à la contrainte temps-réel de la diffusion de contenus multimédia, nous enrichissons nos modèles avec la donnée temporelle. La donnée temporelle permet d'effectuer des *handoffs* en temps requis. En effet, déterminer la position future du terminal mobile ne suffit pas. Il faut également que le handoff soit effectué suffisamment tôt pour être utile, et

suffisamment tard pour ne pas occuper le réseau inutilement. Les expérimentations que nous avons menées sur nos modèles nous permettront de juger de l'efficacité des modèles proposés.

Enfin, une partie en annexes présente les notions fondamentales des domaines de la géolocalisation et de la modélisation de données applicables à la prédiction de la mobilité. En particulier, les unités utilisées pour quantifier la puissance d'un signal radio, support de 802.11, et les phénomènes qui y sont appliqués, sont présentés. La norme 802.11 et ses déclinaisons y sont également décrites afin d'en connaître le fonctionnement. Enfin, les modèles de Markov et les réseaux bayésiens sont exposés car ils seront employés dans la prédiction de la mobilité. Ces notions sont requises pour la compréhension du mémoire car nous fondons nos travaux sur elles.

Fournir des services dépendants du contexte et dans une situation de mobilité requiert que plusieurs conditions soient remplies. Concernant la dépendance au contexte, le terminal mobile doit pouvoir être géolocalisé, c'est-à-dire qu'on peut déterminer sa position dans l'espace. Alors que le GPS donne une position à l'extérieur des bâtiments, il est nécessaire de concevoir un système de géolocalisation à l'intérieur pour fournir des services dépendants du contexte quelle que soit la position de l'utilisateur. Concernant l'existence des services en mobilité, il est nécessaire de mettre en œuvre une procédure nommée handoff.

Nous nous plaçons dans le contexte des réseaux de norme 802.11. En effet, cette norme est largement répandue et bon marché. Elle offre de plus un débit élevé (54 Mbps pour du 802.11g). Les réseaux 802.11 sont souvent déployés à l'intérieur des bâtiments ou en périphérie de ceux-ci, là où le GPS est absent ou moins précis. Bien que nos travaux sont initiés sur les réseaux 802.11, nous n'excluons pas leur adaptation partielle à d'autres supports, tels que 802.15 ou la transmission en bande ultra large (*Ultra Wide Band*).

Nous proposons une approche proactive du handoff, basée sur la prédiction de la mobilité des terminaux mobiles. En effet, prévoir la localisation probable d'un terminal mobile dans un futur proche permet de préparer la procédure de handoff en la facilitant pour la rendre plus transparente pour les applications et pour l'utilisateur. La prédiction de la mobilité permet également de choisir judicieusement les antennes utilisées pour minimiser globalement les *handoffs*.

Première partie

Géolocalisation de terminaux mobiles dans les réseaux Wi-Fi

Chapitre 1

État de l’art de la géolocalisation dans les réseaux Wi-Fi

Introduction

La **géolocalisation**, indispensable pour fournir des services dépendant du contexte d’un terminal mobile et requise pour la prédiction de la mobilité, consiste à détecter la position de l’individu via son terminal mobile. La géolocalisation consiste à déterminer les coordonnées d’un terminal mobile. Les coordonnées peuvent être relatives, c’est-à-dire définies par un repère d’origine arbitraire. Elles peuvent sinon être géographiques, c’est-à-dire par rapport au centre de la Terre, en coordonnées polaires.

La géolocalisation à l’extérieur est réalisée par des systèmes de géolocalisation par satellite (*Global Navigation Satellite System*, GNSS). Cependant, dans des environnements couverts, les forêts et les villes en particulier, les GNSS fonctionnent mal, amenant des erreurs de plusieurs dizaines de mètres, voire pas du tout. Ils ne fonctionnent également pas à l’intérieur des bâtiments, sauf avec l’usage de répéteurs [2]. C’est pourquoi des alternatives basées sur des appareils facilement accessibles et bon marché voient le jour. En particulier, l’utilisation de la norme IEEE 802.11 est intéressante pour le développement d’un système de géolocalisation en intérieur. En effet, les appareils Wi-Fi (compatibles IEEE 802.11) sont déployés en grand nombre tant chez les particuliers qu’en entreprise. De plus, le coût de ces appareils est de quelques dizaines d’euros seulement. Ils ont également un débit intéressant, permettant de coupler la fonction de localisation et la fourniture des services au sein d’un réseau unique. Ces atouts font des réseaux Wi-Fi le support idéal à nos travaux.

Dans un premier temps, nous présentons les manières de classifier les systèmes de géolocalisation. Dans un second temps, nous présentons les systèmes de géolocalisation basés sur des appareils dédiés au positionnement. Bien qu’elles ne soient pas basées sur le Wi-Fi, les techniques employées, tant dans le calcul que l’exploitation, sont susceptibles d’être appliquées aux réseaux Wi-Fi. Troisièmement, en nous appuyant sur la classification définie dans la première section, nous décrivons les travaux de recherche dans le domaine de la géolocalisation de terminaux mobiles de la norme IEEE 802.11. Quatrièmement, nous analysons les travaux relatifs à la norme IEEE 802.11 en mettant en exergue leurs défauts et qualités par rapport à notre problématique de système de géolocalisation en intérieur facile à déployer.

1.1 Terminologie de l'infrastructure des systèmes de géolocalisation

Dans cette section, les termes employés dans ce chapitre sont précisés ou restreints à la problématique de la géolocalisation.

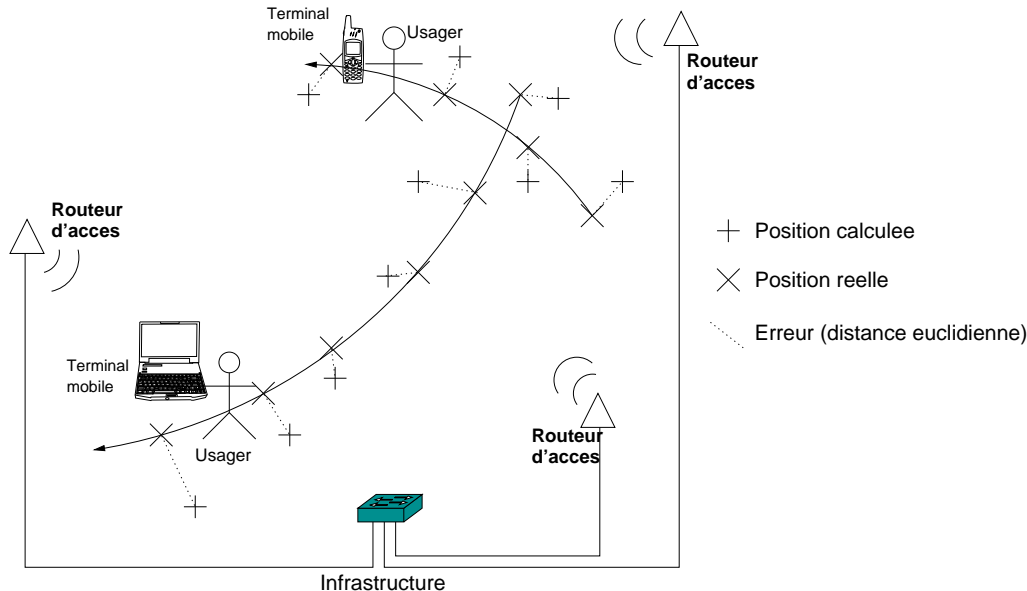


FIG. 1.1 – Illustration des définitions et rôles dans la géolocalisation.

Nous appelons *terminal mobile* un équipement du réseau capable de se mouvoir, généralement porté par une personne. Le terminal mobile doit disposer d'une capacité de calcul, d'une mémoire et d'une connectivité réseau sans-fil. On considère en particulier que les ordinateurs portables, les téléphones mobiles et les assistants personnels digitaux (*Personal Digital Assistant* ou PDA) sont des terminaux mobiles. Un terminal mobile peut être statique à un instant donné, mais sa connexion sans-fil lui permet potentiellement d'être mobile.

L'*usager* ou l'*utilisateur* est la personne qui tient le terminal mobile et utilise les services disponibles sur le réseau. En particulier, l'usager requiert le service de géolocalisation.

Le *réseau câblé* ou *réseau d'infrastructure* est la partie du réseau dans laquelle les appareils sont reliés entre eux par des liaisons filaires. C'est généralement le cœur du réseau, dans lequel on trouve les machines fournissant les services comme celui de la diffusion de contenus média riches.

La *partie mobile* du réseau est l'ensemble des terminaux mobiles reliés au réseau par une liaison sans-fil. Nous considérons que les terminaux mobiles sont clients de services offerts par des serveurs situés dans le réseau câblé.

Le *routeur d'accès* est l'appareil de l'infrastructure du réseau qui assure le lien entre l'infrastructure du réseau câblé et sa partie mobile. Les routeurs d'accès sont, par exemple, les stations de base (*Base Transmission Station* ou BTS) de la téléphonie mobile ou les points d'accès (*Access Points* ou AP) des réseaux Wi-Fi.

Une *méthode de géolocalisation* est l'ensemble des traitements qui permettent de déterminer la position d'un terminal mobile à partir de données quantifiables comme la puissance des signaux reçus. Le *système de géolocalisation* est l'ensemble logiciel et matériel qui met en œuvre une méthode de géolocalisation. Un système de géolocalisation produit une *estimation de la position* d'un terminal mobile. On parlera également de système de localisation pour s'y référer.

Pour quantifier la *précision* d'un système de géolocalisation, nous considérons l'erreur du positionnement qu'il effectue. L'erreur est la distance euclidienne entre les coordonnées de la position calculée par le système de géolocalisation et les coordonnées de la position réelle du terminal mobile. Plus cette distance est grande, moins le système de géolocalisation est précis.

La figure 1.1 met en situation les notions définies. Deux usagers se déplacent avec leurs terminaux mobiles. Les routeurs d'accès au réseau d'infrastructure sont connectés par un commutateur. Les croix en forme de 'X' sont les positions réelles de chacun des terminaux mobiles au cours de leurs déplacements. Les croix en forme de '+' sont les positions calculées par le système de géolocalisation. L'erreur, évaluée par la distance euclidienne entre les vraies positions et les positions estimées, est représentée par les lignes pointillées.

1.2 Classification des systèmes de géolocalisation

Dans cette section, nous proposons dans un premier temps des critères de classification des systèmes de géolocalisation. La classification est nécessaire pour regrouper et analyser des méthodes de résolution semblables et en extraire les grandes lignes. Dans un second temps, nous énumérons les critères d'évaluation des performances des systèmes de géolocalisation que nous étudions. Ces critères permettent de juger de la précision mais également de la pertinence des systèmes étudiés dans le cadre de nos travaux.

1.2.1 Critères liés aux techniques et méthodes de géolocalisation

Établir une classification des systèmes de géolocalisation est important pour appréhender les différentes méthodes mises en œuvre dans les systèmes de géolocalisation étudiés. Il est ainsi plus facile d'étudier les systèmes qui nous intéressent et de se baser sur ceux qui semblent les plus pertinents dans le cadre de la géolocalisation de terminaux Wi-Fi en intérieur.

Nous classons d'une part les systèmes de géolocalisation selon qu'ils reposent sur un dispositif dédié ou sur la norme IEEE 802.11. D'autre part, chaque famille se voit décomposée elle-même selon qu'elle localise le terminal mobile dans un ensemble infini indénombrable (continu) ou fini et dénombrable (discret). La géolocalisation discrète s'appuie sur la connexion du terminal mobile à un équipement du réseau d'infrastructure. La géolocalisation continue s'appuie sur la trilatération.

Au sein de la catégorie des systèmes de géolocalisation basés sur des appareils Wi-Fi, nous distinguons également une sous-famille de systèmes de géolocalisation discrets et une sous-famille de systèmes de géolocalisation continus. Les systèmes de géolocalisation discrets basés sur Wi-Fi fonctionnent par comparaison d'empreintes des puissances des signaux reçus avec une base de données de mesures préalables. On appelle cette base de données la **cartographie des puissances**, qui est composée de **données de référence** aussi appelées **points de référence**. Les systèmes de géolocalisation continus basés sur Wi-Fi fonctionnent en calculant la distance

entre des points connus, typiquement les points d'accès, et le terminal mobile. Puis, ils calculent la position du terminal mobile par trilatération. Nous nous référons également aux modèles basés sur une cartographie des puissances par le terme **modèle de référencement**.

Un dernier critère de classification est restreint aux systèmes de géolocalisation Wi-Fi discrets. Il s'agit du déterminisme de la localisation. Ce critère définit si le système s'appuie sur une représentation simple des données de référence. En particulier, la moyenne des puissances mesurées en un point est une représentation simple. Dans ce cas, le système est **déterministe**. Quand le système s'appuie sur une représentation statistique, il associe à chaque point de référence une distribution aléatoire dont les paramètres sont basés sur les mesures effectuées. On dit alors que le système est probabiliste.

1.2.2 Symbologie

Il existe plusieurs propriétés permettant de juger une technique de géolocalisation. Dans un article traitant des systèmes de géolocalisation en intérieur[3], J. A. Tauber détermine plusieurs propriétés définissant les systèmes de géolocalisation :

- ▶ la symbologie ;
- ▶ les erreurs ;
- ▶ la fréquence de localisation ;
- ▶ le délai de localisation ;
- ▶ l'extensibilité ;
- ▶ le coût ;
- ▶ le choix d'un calcul centralisé ou distribué.

La **symbologie** du système est la caractéristique fondamentale d'un système de géolocalisation. Elle consiste à définir la relation entre l'espace en 3 dimensions dans lequel nous évoluons et sa représentation par le système de géolocalisation. Cette représentation est, par exemple, les coordonnées cartésiennes ou des informations liées à la topologie du bâtiment. De manière générale, la symbologie du système de géolocalisation est déterminée par l'emploi qui en sera fait. Les coordonnées cartésiennes sont particulières car elles incluent toute symbologie basée sur un découpage de l'espace. On peut y adjoindre des informations indiquant les services disponibles selon la localisation du terminal mobile.

L'**erreur** générée par le système de géolocalisation est un critère d'analyse important. Elle permet de juger si le système de géolocalisation est pertinent pour l'usage qui lui sera dévolu. La localisation étant basée sur des mesures des propriétés physiques du médium de transmission, les erreurs sont inévitables. Les erreurs peuvent être détectées et corrigées, lorsqu'elles sont systématiques. Elles peuvent sinon être traitées par des méthodes statistiques lorsqu'on ne connaît pas les conditions de leur apparition.

La **fréquence de localisation** est la vitesse à laquelle les positions sont calculées. Ce critère prend son importance quand un système de géolocalisation centralisé doit servir de nombreux clients. Il sera également déterminant pour l'extensibilité du système devant l'augmentation du nombre de terminaux mobiles. Le **délai de localisation** est le temps qui sépare le déplacement du terminal mobile de sa localisation par le système. Le délai de localisation est important car il

détermine l'erreur systématique du système de localisation, liée au déplacement du mobile entre l'instant de requête de sa position et l'instant de réception de la réponse. Cette propriété est importante dans les applications de réalité virtuelle.

L'**extensibilité du système** est l'aptitude du système à être étendu soit à une zone de déploiement plus grande, soit à un plus grand nombre de terminaux mobiles, voire aux deux. Les tailles de la zone de déploiement ou de la flotte de terminaux mobiles localisés sont à comparer à celles pour lesquelles le système de géolocalisation a été développé. En effet, plus on étend le système de géolocalisation à des zones vastes ou à une quantité importante de terminaux mobiles, plus la quantité requise de calculs croît.

Le **coût** est défini par plusieurs paramètres. Le coût financier dépend des salaires des personnels qui déploient et maintiennent le système de géolocalisation et du prix des équipements à déployer pour le fonctionnement du système de géolocalisation. Le coût énergétique est dépendant de la consommation du système de géolocalisation. Il est important en ce qui concerne les terminaux mobiles, car il affecte leur autonomie. Le coût en temps inclut le temps d'installation et les temps de maintenance.

J. A. Tauber souligne aussi la différence entre le choix d'un développement **centralisé ou distribué**. Dans un système centralisé, la position est calculée par un serveur. Dans un système distribué, la position est calculée soit par plusieurs serveurs soit par les terminaux eux-mêmes. Ce choix technique a un impact sur l'extensibilité du système. Le choix de l'architecture peut être dicté par le fonctionnement du système de géolocalisation ou par les capacités des différents appareils déployés.

1.3 Géolocalisation discrète

Dans cette partie, nous présentons les techniques de géolocalisation basées, soit sur une approximation de la position suivant un critère prédéfini, soit sur la collecte des données mesurées à diverses positions géographiques du terrain de mise en œuvre. Ces techniques sont qualifiées de discrètes car l'ensemble des positions possibles est restreint à un sous-ensemble de points de l'espace cartésien. Ce sous-ensemble est l'ensemble des points auxquels des relevés de mesures ont été effectués.

1.3.1 Techniques de géolocalisation cellulaire

Les techniques de localisation cellulaires définissent comme approximation de la position d'un terminal mobile son antenne de rattachement à l'infrastructure du réseau (BTS, point d'accès Wi-Fi [4], etc.). La précision de telles méthodes de positionnement est fonction de la portée des équipements d'infrastructure. Par exemple, le cas du GSM permet une précision de l'ordre de la centaine de mètres (en environnement urbain) à plusieurs kilomètres (environnement rural). Par opposition, un tel système basé sur la norme 802.11 aurait une précision de l'ordre de quelques dizaines de mètres, grâce à sa portée plus courte. Ce type de méthode de géolocalisation est utile pour acheminer un service au client mobile. Par exemple, l'acheminement d'un appel téléphonique vers un mobile requiert de savoir à quelle antenne (BTS) le mobile est connecté.

1.3.2 Communications courte portée

L'utilisation de terminaux disposant d'une faible portée permet la localisation. La méthode est la même que pour les techniques cellulaires, mais on ne parle plus de cellule au vu des technologies employées. Les dispositifs de faible portée utilisés sont les infrarouges, les RFID (*Radio Frequency IDentification*) et bluetooth. En se basant sur les dispositifs à sa portée, la position d'un terminal mobile est facile à estimer. La précision est acceptable pour beaucoup d'applications, en particulier avec RFID et les infrarouges. Bluetooth ayant une plus grande portée, la précision est moindre mais elle reste meilleure que dans les techniques de géolocalisation cellulaires.

Active Badge [5] est un système de géolocalisation basé sur l'utilisation des infrarouges (IR). Le système d'Active Badge est décomposé en deux parties. D'une part, des badges contenant un émetteur IR sont portés par les utilisateurs, d'autre part, un réseau de récepteurs est installé dans le bâtiment. Les badges émettent une balise d'une durée d'un dixième de seconde toutes les 15 secondes. La balise est un code unique associé à un utilisateur. Le réseau de récepteurs reçoit les signaux et en déduit la localisation de l'utilisateur. Grâce à la réflexion du signal IR et l'impossibilité de traverser les murs, il est rare que plusieurs récepteurs reçoivent une balise identique. En cas de réceptions multiples, des algorithmes effectuent un choix parmi les réponses potentielles. Le badge porté par les utilisateurs pèse 40 g pour des dimensions de $55 \times 55 \times 7$ mm.

CarpetLAN [6], propose une approche singulière de l'utilisation des capteurs : ceux-ci servent de support à la fois à la localisation et au réseau de communication. CarpetLAN est architecturé autour d'un tapis de capteurs à la surface du sol. Un faible courant porteur traverse ces capteurs, et rejoint le terminal mobile en passant au travers du corps humain ou du mobilier ayant la particularité d'être conducteur. Ce courant porteur autorise les échanges de données à une vitesse de 10 Mbps. Ainsi, une personne tenant un assistant personnel peut marcher pour se déplacer, tout en gardant une connexion au réseau. Le rattachement d'un dispositif à un capteur permet de géolocaliser l'appareil. De ce point de vue, CarpetLAN utilise une technique de géolocalisation cellulaire. La technologie de CarpetLAN permet une précision de positionnement pratique de l'ordre du mètre. Cependant son coût est très élevé.

1.3.3 Usage des ultrasons

Le système Cricket [7] calcule les distances en utilisant la différence des temps de réception entre des signaux radio et ultrasons synchronisés. Chaque balise émet un signal radio identifiant de façon unique sa position. Les terminaux mobiles calculent les distances entre eux et les balises reçues. La distance la plus courte correspond à la position de la balise la plus proche. Le terminal déduit alors sa position comme étant celle de la balise la plus proche. Afin de supprimer les ambiguïtés dans l'association entre un signal radio et un signal ultrason, deux algorithmes sont utilisés. Une opération préalable et commune aux deux algorithmes est le tri des distances. Pour chaque balise, on calcule sa distance par rapport au terminal mobile selon chaque signal ultrason. Les distances sont arrondies aux 25 cm supérieurs. Puis, la fréquence de chaque distance est calculée et ordonnée par ordre croissant. Le premier algorithme, *MinMean*, choisit la balise qui a une moyenne des distances la plus faible. Le second algorithme, *MinMode*,

sélectionne la balise dont le mode¹ de la distribution des distances est le plus petit. Lors des tests, les terminaux identifient correctement la pièce dans laquelle ils sont dans 95 % des cas lorsqu'ils sont statiques. Des résultats similaires ont été obtenus avec des terminaux mobiles, l'algorithme *Minmode* et une fenêtre de signaux de 5 éléments.

WALRUS [8] est une autre forme de localisation basée sur l'usage combiné des ultrasons et d'ondes radio. Les signaux ultrason sont émis par les haut-parleurs des PC localisés dans les bureaux du bâtiment. Les récepteurs sont des microphones intégrés aux terminaux mobiles². Les PC qui émettent les signaux ultrason et radio sont les serveurs de balises (*Beacon Server*). Les serveurs de balise émettent simultanément une trame radio 802.11 et une impulsion ultrason. Le signal ultrason ne contient pas de donnée, il s'agit seulement d'une impulsion. L'objectif des terminaux mobiles est donc de faire correspondre un signal radio avec au plus un signal ultrason pour déterminer sa position. En effet, les signaux Wi-Fi portent plus loin que les impulsions ultrason émises par les serveurs de balise. Chaque trame 802.11 contient le label de la salle d'où il est émis. Le terminal mobile conserve une liste des trames 802.11 reçues. Le paramètre MDT (*Maximum Detection Time* ou temps maximal de détection) indique le temps maximal d'attente après réception d'une trame 802.11 durant lequel une impulsion ultrason peut être reçue.

1.4 Géolocalisation continue

La géolocalisation continue permet de déterminer la position d'un terminal mobile dans l'espace cartésien. Cette famille de méthodes de positionnement s'appuie sur le calcul mathématique de la position du terminal mobile, en particulier à l'aide de la trilatération. Nous présentons d'abord les systèmes les plus connus : les systèmes de géolocalisation par satellite. Puis, nous décrivons des systèmes de géolocalisation dédiés à un usage à l'intérieur des bâtiments.

1.4.1 Positionnement par satellite

L'exemple le plus connu de positionnement par satellite est le GPS (Global Positioning System) initialement baptisé Navstar [9]. Il utilise une constellation de 24 satellites en orbite basse à 20200 km d'altitude (les émetteurs) et des récepteurs GPS dans les véhicules à positionner. Il est nécessaire que les récepteurs et les satellites soient synchronisés car la précision de positionnement du GPS dépend des horloges des dispositifs. À intervalle régulier, les satellites émettent des trames. Le processus de positionnement est basé sur la connaissance de la vitesse des ondes émises par les satellites. Ceux-ci émettent à intervalle régulier un signal contenant leur identifiant et la date d'émission. Du fait de leur synchronisation avec les capteurs GPS, ces derniers peuvent déterminer par le calcul de $d = V.t$, leur distance d par rapport à chaque satellite. La vitesse V est connue et égale à c ($3.10^8 m.s^{-1}$) et le temps t est déterminé par différence de la date de réception du signal avec l'estampille du signal. La trilatération permet de connaître la position du capteur GPS. Au minimum, trois satellites doivent être captés pour obtenir une position. L'onde radio pouvant subir plusieurs phénomènes qui la ralentiront, ainsi que la dérive des horloges entre les satellites et les capteurs, la position calculée est soumise à une erreur. Un quatrième satellite, s'il est disponible, permet d'accroître la précision en supprimant

¹Valeur la plus représentée de l'échantillon.

²Comme c'est le cas pour une majorité des PDA et ordinateurs portables

l'erreur commise. Cette technique est le *pseudorangeing*. D'autres satellites amènent également plus de précision, mais de façon moins perceptible. Le GPS différentiel [9] (*Differential GPS* ou DGPS) utilise la différence de temps de réception entre le récepteur et un récepteur proche dont la position a déjà été déterminée. Le GPS nécessitant une ligne de vue entre les satellites et le récepteur GPS, il n'est pas adapté à l'usage dans les bâtiments. Des répéteurs GPS [10] permettent cependant de palier ce problème.

Le GPS a deux concurrents. GLONASS (*Global'naya Navigatsionnaya Sputnikovaya Sistema*) qui signifie Système Global de Navigation par Satellite est l'alternative russe du GPS. Glonass comporte 24 satellites à une altitude de 19100 km. GALILEO est le futur projet européen de système de positionnement par satellite. Il est en test depuis 2004, sera opérationnel en 2008 et complètement achevé en 2010. Il est prévu de déployer 30 satellites à une altitude de 23616 km.

1.4.2 Utilisation de capteurs

Le système active bat [11] utilise la combinaison des ultrasons et des ondes radio. Des émetteurs, dont les coordonnées sont connues, sont répartis dans la zone de mise en œuvre. Chaque émetteur émet un signal radio et un signal ultrasonore simultanés. Les récepteurs ont deux antennes séparées de plusieurs centimètres. Quand le récepteur reçoit le signal radio, il mesure le temps mis par le signal ultrasonore pour parcourir la distance entre l'émetteur et chacune des antennes. En effet, le temps de propagation de l'onde radio (vitesse de la lumière) est considéré comme négligeable par rapport à celui du son. La connaissance de la distance entre l'émetteur et les antennes et de la distance entre les antennes permet de calculer l'orientation du mobile relativement à l'émetteur. La précision de l'angle déterminé est de l'ordre de 3°, c'est-à-dire une erreur de quelques centimètres sur la distance. En utilisant les angles calculés par rapport à plusieurs émetteurs, on peut déterminer la position du terminal par triangulation. En utilisant les distances, on peut la déterminer par trilatération.

Il existe plusieurs types de capteurs permettant de détecter et quantifier les mouvements d'un objet. Parmi ceux-ci, on trouve les accéléromètres, les compas et les gyroscopes. Des systèmes bâtis sur ces éléments permettent de se localiser. Dans le projet PNM (*Pedestrian Navigation Module*) [12], les inconvénients des systèmes classiques sont contrebalancés par leur combinaison. Le PNM combine un gyroscope, un compas magnétique et un accéléromètre tri-axial. La perturbation du compas est palliée par le gyroscope, qui est lui même recalé par l'accéléromètre lorsqu'il dérive. La précision finale de cet appareil est de 5 % de la distance parcourue. L'erreur étant due à une dérive des capteurs, elle est cumulative et croît avec la distance parcourue. L'inconvénient majeur de cet appareil est sa disponibilité.

Il est aussi possible de combiner les résultats obtenus par plusieurs capteurs. Une méthode est la fusion de capteurs (*Sensors Fusion*). Dans *SELFLOC* [13], Y. Gwon et al. proposent de combiner les résultats de plusieurs systèmes de positionnement. La fusion de capteurs consiste à combiner des résultats issus d'approches différentes. Un poids est attribué à chaque méthode. La première phase est l'apprentissage des poids attribués. Il est basé sur la présentation de données d'entrée au modèle, ainsi que des résultats. En comparant le résultat du modèle et le résultat réel, connu dans les données d'apprentissage, on minimise la différence en modifiant les poids de chaque méthode. SELFLOC s'applique tant à des supports divers, comme un système basé sur

bluetooth d'une part et Wi-Fi d'autre part, qu'à des résultats d'algorithmes différents basés sur un support identique. SELFLOC permet d'atteindre une précision de l'ordre de 1,5 mètres en combinant les résultats de la localisation par rapport à 2 points d'accès et 2 appareils bluetooth. Ses auteurs font également une comparaison avec *Ekahau* [14]. Le résultat du système Ekahau est de 4 mètres seul. Il est de 2 mètres lorsqu'on l'intègre à SELFLOC et les résultats de 3 appareils bluetooth.

Les ondes radio ne sont pas les seuls supports à la géolocalisation de terminaux mobiles. Une onde sonore peut également être exploitée dans le cadre de la géolocalisation. *Cricket Compass* est une extension à *Cricket* qui détermine l'orientation du terminal mobile. Pour y parvenir, la différence de distance entre deux extrémités du dispositif est utilisée. Ce système obtient une précision de l'ordre du centimètre.

La mesure de la puissance du signal peut également s'utiliser pour déterminer la distance entre un émetteur et un récepteur. En effet, la puissance du signal émis diminue avec l'augmentation de la distance parcourue. Ainsi, plus le récepteur est loin de l'émetteur du signal, plus la puissance du signal qu'il recevra sera faible. D'autres paramètres entrent en compte dans l'affaiblissement du signal, en particulier dans les environnements intérieurs comme les bâtiments, mais la distance est un facteur majeur d'affaiblissement du signal. La mesure de ce dernier est donc un bon moyen d'estimer la distance entre un récepteur et l'émetteur du signal mesuré.

1.5 Géolocalisation des terminaux Wi-Fi

Dans cette section, nous nous intéressons aux systèmes de géolocalisation de terminaux Wi-Fi. Après une présentation de la norme IEEE 802.11, définissant les réseaux de type Wi-Fi, nous exposons les classifications possibles de ces systèmes. Puis, nous présentons les techniques de géolocalisation basées sur la collecte de mesures des caractéristiques des signaux reçus. Ensuite, nous décrivons les techniques de géolocalisation basées sur la modélisation de la trajectoire des ondes radio, porteuses des transmissions Wi-Fi. Enfin, nous présentons des techniques basées sur des modèles d'apprentissage, tels que les réseaux de neurones et les réseaux bayésiens.

1.5.1 Collecte de mesures des caractéristiques des signaux reçus

La collecte de données établit une base de données des observations physiques du réseau sans-fil, par des mesures de puissance du signal, de rapport signal sur bruit ou tout autre grandeur mesurable permettant de quantifier le signal. On qualifie également cette famille d'ensemble de techniques basées sur une cartographie des puissances. Le projet fondamental basé sur la collecte de données a été initié par Microsoft Research. RADAR, conçu par Bahl et al. [15], est basé sur l'enregistrement de mesures de puissance des signaux reçus. Une base de données contient les mesures en relation avec leur position. Une mesure et sa localisation forment un point de référence. Par comparaison de nouvelles valeurs mesurées avec le contenu de la base de données, on peut déterminer la localisation d'un terminal mobile. Alors que RADAR propose une comparaison déterministe, d'autres approches sont, elles, basées sur des méthodes probabilistes. Par ailleurs, la base de données peut être constituée d'une seconde manière : il est possible, avec un bon modèle de propagation des ondes, de calculer pour chaque point de référence les puissances des signaux reçus.

Approche déterministe

L'approche déterministe constitue la base de données de référencement avec une information de puissance moyenne. La géolocalisation est ensuite effectuée par comparaison des mesures en temps réel avec les informations contenues dans la base de données. Tous les systèmes de géolocalisation déterministes stockent au moins les coordonnées physiques et les moyennes des mesures de puissance pour chaque point d'accès du système. On obtient donc une base de données composée d'une suite d'enregistrements relatifs chacun à un point de référence. Un enregistrement est de la forme (p, ss_1, \dots, ss_N) où p sont les coordonnées géographiques du point et le n -uplet (ss_1, \dots, ss_N) est l'ensemble des moyennes de puissance mesurées pour les N points d'accès du système. La base de données peut également contenir d'autres informations relatives à la qualité du signal ou au positionnement des terminaux mobiles. Ainsi, dans le projet RADAR [15], l'orientation du terminal mobile est également stockée dans la base de données.

Bahl et al. observent que le choix du point le plus proche d'une mesure dans l'espace des puissances soulève un problème : une même mesure peut être proche, dans l'espace des puissances, de deux points qui seront éloignés dans l'espace à trois dimensions dans lequel nous nous représentons. Considérant les fluctuations inévitables du signal, il est alors probable que la localisation donnée par RADAR "saute" d'un point à un autre. Cependant, il est invraisemblable qu'un terminal mobile se déplace sur de grandes distances lors d'intervalles de temps courts. Bahl et al. proposent de traduire cette propriété par un algorithme **Viterbi-like** [16].

L'algorithme proposé fonctionne dans le cas du suivi continu du terminal mobile. À chaque mesure de la puissance des signaux reçus par le terminal mobile, les k plus proches voisins de la mesure dans l'espace des puissances sont déterminés. Ces k plus proches voisins sont les k positions les plus vraisemblables du mobile, considérant la mesure courante. Un historique de n ensembles de k plus proches voisins est conservé. On peut voir ces n ensembles de k points comme représentés dans la figure 1.2.

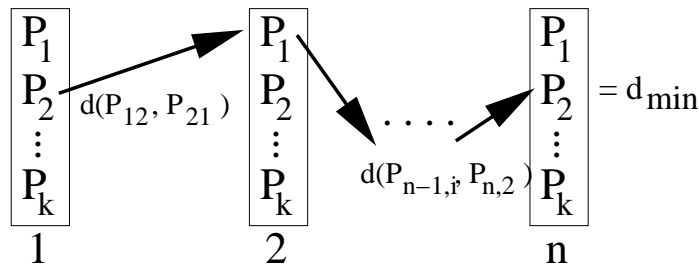


FIG. 1.2 – Algorithme Viterbi-like dans le système RADAR.

Dans la figure 1.2, les arcs (représentés par les flèches) existent uniquement entre des points appartenant à des ensembles consécutifs, c'est-à-dire les ensembles construits à des itérations consécutives. Les arcs sont pondérés par les distances entre les points qu'ils connectent. Tous les chemins possibles sont testés. Un chemin commence par un point du premier ensemble et finit par un point du n -ième ensemble. Il passe par un point de chaque ensemble, dans l'ordre de leur construction (2^{nd} puis 3^{ieme} , etc.). Le poids total d'un chemin est la somme des distances portées par les arcs du chemin. Plus le poids est important, plus la distance parcourue est grande et moins le chemin est considéré comme vraisemblable. Le plus court chemin est considéré comme

	Sans profil	Avec profils
Erreur moyenne	5,9	2
Erreur médiane	5,1	1,9
Erreur à 90%	11	3

FIG. 1.3 – Résultats du système RADAR (erreurs en mètres).

le chemin le plus vraisemblable et il est alors sélectionné. Le point de départ de ce chemin est renvoyé comme étant la position du terminal mobile. Cette méthode induit un délai dans le calcul de la position car il nécessite de constituer $n - 1$ ensembles de points supplémentaires, donc de procéder à autant de mesures de la puissance de signaux reçus. L'algorithme Viterbi-like présenté par Bahl et al. ne tient pas compte de la topologie des bâtiments. En effet, l'utilisation de la distance euclidienne conduit souvent à sous-estimer la distance réelle parcourue par le terminal mobile entre deux points : la présence d'obstacles, les murs en particulier, oblige le porteur d'un terminal mobile à faire des détours. Enfin, la complexité de l'algorithme Viterbi-like est exponentielle : il est nécessaire d'étudier tous les chemins possibles passant par 1 point de chaque ensemble constitué à chaque itération. Considérant les paramètres n et k , où n est le nombre d'ensembles conservés et k le nombre de points par ensemble, il existe k^n chemins possibles, soit k^n sommes de distances à effectuer. Ce grand nombre d'opérations est effectué à chaque demande de localisation par le terminal mobile.

La géolocalisation d'un terminal mobile est effectuée par une comparaison d'une mesure courante avec le contenu de la base de données. On qualifie de mesure courante l'ensemble des couples (P, id) mesurés dans la phase de géolocalisation. P est une puissance de signal reçu et id est l'identifiant du point d'accès correspondant. La comparaison est faite par calcul de la distance euclidienne dans l'espace du signal. L'espace du signal est un espace à N dimensions où N est le nombre de points d'accès du système de géolocalisation. La distance calculée est celle entre la mesure courante et les mesures enregistrées. L'enregistrement pour lequel la distance est minimal est élu comme solution. C'est l'algorithme le plus simple pour déterminer la position. Il est notamment employé dans le système de géolocalisation de la *Carnegie Mellon University* (CMU) sous le nom d'algorithme de *Pattern Matching* (CMU-PM) [17].

Bahl et al. remarquent que la puissance du signal varie différemment selon les heures. En particulier, aux heures de bureau, la puissance moyenne varie beaucoup et est inférieure à la puissance moyenne le reste du temps. Ceci est dû aux personnes présentes dans le bâtiment, qui perturbent le signal. Pour palier ce phénomène, un système de profils est mis en place. Deux profils sont utilisés : un profil de jour, adapté à une forte fréquentation, et un profil de nuit. Les points d'accès changent dynamiquement de profil en mesurant les puissances de leurs signaux réciproques. Le tableau 1.3 indique les résultats obtenus par le système RADAR en fonction de l'utilisation des profils.

Cette approche est améliorée par M. Brunato et C. K. Kalló dans leur système *Transparent Location Fingerprinting* [18]. Les auteurs de ce système déterminent la position du terminal mobile en sélectionnant plusieurs points de la base de données de référencement. En considérant

k points sélectionnés, le point final est calculé comme suit :

$$p = \frac{\sum_{j=1}^k \frac{1}{d(ss_{ij};ss)+\varepsilon} \cdot p_{ij}}{\sum_{j=1}^k \frac{1}{d(ss_{ij};ss)+\varepsilon}}$$

où $d(ss_{ij};ss)$ est la distance euclidienne entre les mesures ss_{ij} de la base de données et la mesure courante ss , ε est une petite constante pour éviter une division par zéro et p ainsi que p_i sont respectivement les coordonnées du point recherché et des points de la base de données. Ce calcul détermine un point dont les coordonnées sont les moyennes pondérées des k points élus. La pondération est dépendante de l'inverse de la distance de chaque point de référence par rapport à la mesure courante dans l'espace des puissances. Cette approche est dérivée de la méthode des *k plus proches* qui sélectionne k points pour en faire la moyenne. L'erreur moyenne annoncée par les auteurs de ce système est de 1,78 mètres.

Précalcul de la cartographie des puissances

Mesurer les puissances reçues en différents points de référence est une possibilité pour créer la cartographie des puissances. Cependant, cette méthode requiert de se déplacer avec un terminal mobile pour se positionner aux coordonnées de chaque point et y procéder à la mesure. À l'inverse, si l'on peut disposer d'une description précise du bâtiment de mise en œuvre, il est possible de calculer la cartographie des puissances. Pour cela, il faut un modèle précis de la propagation des ondes en intérieur.

Les auteurs de RADAR proposent le calcul basé sur les travaux de Seidel et al. [19]. La formule utilisée pour calculer la puissance du signal est :

$$P_{dBm}(d) = P_{dBm}(d_0) - 10n \log\left(\frac{d}{d_0}\right) - \alpha \times WEF$$

où $\alpha = C$ si $nW \geq C$ et $\alpha = nW$ sinon. WEF est le *Wall Effect Factor* qui détermine l'impact des murs sur la puissance du signal. Les distances d et d_0 sont respectivement la distance entre l'émetteur et le récepteur et une distance de référence. La valeur n détermine la vitesse à laquelle le signal s'atténue avec la distance. Le nombre de murs nW permet de comptabiliser l'atténuation du signal due aux murs traversés. La valeur de nW est limitée à un seuil C au delà duquel l'atténuation se fait moins sentir. WEF est une valeur en dBm.

K. Runser et al. proposent une autre méthode du calcul de la puissance du signal. Il s'agit de MR-FDPF (*Multi-Resolution Fourier Domain ParFlow*) [20]. Cette méthode a pour origine le calcul de la couverture des BTS dans les réseaux GSM à des fins de planification. C'est une méthode discrète, qui consiste à découper la zone de calcul en une grille. Le champ électrique, qui forme l'onde radio, est calculé à l'aide de 5 composantes, les flux, propagés de bloc en bloc dans la grille. Des paramètres sont à déterminer pour connaître les propriétés des matériaux. Ils sont déterminés expérimentalement par mesure.

Approche probabiliste

Les approches probabilistes consistent à utiliser un ensemble de valeurs de SS pour chaque point de référence. Les valeurs sont représentées par une distribution. Généralement, une distribution suivant une loi normale est choisie. L'idée de l'utilisation de plusieurs mesures est d'absorber les variations inévitables des mesures par la quantité. De cette façon, les mesures utilisées sont plus fiables car la redondance des échantillons de mesures réduit les effets des interférences en produisant un ensemble de mesures réparties selon leurs probabilités d'être observées.

Ekahau [14] est un système de géolocalisation commercialisé³. Ses auteurs considèrent la localisation comme un problème d'apprentissage. Les données de référence sont utilisées pour construire un modèle. Le modèle sous-jacent d'Ekahau est construit grâce à la formule de Bayes. Déterminer la position du terminal mobile revient à calculer :

$$p(l/o) = \frac{p(o/l)p(l)}{p(o)}$$

où :

- ▶ l est une position, contenue dans les points de référence ;
- ▶ o est une observation, c'est-à-dire un ensemble de mesures ;
- ▶ $p(o/l)$ est la probabilité d'observer o sachant que l'on est à la position l . Cette probabilité est déterminée à partir des données de référence ;
- ▶ $p(l)$ est la probabilité *a priori* d'être à la position l ;
- ▶ $p(o) = \sum_{l' \in L} p(o/l')p(l')$ avec L l'ensemble des points de référence.

Estimer $p(o/l)$ est possible en se basant soit sur les histogrammes des puissances, soit sur une représentation gaussienne. La position retournée par le système est l'espérance de la position, soit la moyenne des positions pondérées par leurs probabilités respectives. Les tests du système comparent la méthode simple utilisée dans RADAR, la distribution en histogramme et la distribution gaussienne. Les tests sont effectués avec 1 mesure et 20 mesures, simulant un mobile lent. Les résultats sont décrits dans le tableau 1.4. La précision est en faveur de la distribution gaussienne quand il n'y a qu'une mesure (mobile rapide) et en faveur de la distribution en histogramme quand il y a 20 mesures (mobile lent).

	1 mesure		20 mesures	
Méthode	Moyenne	90%	Moyenne	90%
RADAR	3,71	7,23	1,67	2,8
Gauss	2,57	4,6	1,69	3,07
Histogramme	2,76	5,37	1,56	2,76

FIG. 1.4 – Résultats d'Ekahau (erreurs en mètres).

Le problème auquel on se confronte en utilisant beaucoup de données, en particulier la quantité nécessaire pour une description statistique, est le coût des calculs à appliquer pour

³<http://www.ekahau.com>

traiter toutes les données. M. Youssef et al. proposent un système de géolocalisation, nommé HORUS, pouvant fonctionner suivant deux méthodes de géolocalisation [21] :

- le joint clustering ;
- la triangulation incrémentale (incremental triangulation).

Dans le système HORUS, une distribution de SS par histogrammes est définie pour chaque couple de point de référence et point d'accès. Soit un point d'accès A et un point de référence R . Soit N le nombre de mesures disponibles pour (A, R) et N_V les occurrences de la valeur de puissance du signal V . La probabilité que la valeur de la puissance du signal soit de V est :

$$P(V/A, R) = \frac{N_V}{N}$$

Le joint clustering consiste à identifier dans un premier temps la distribution jointe des valeurs de SS à chaque point de référence. Dans un second temps, il convient de regrouper les points de référence en groupes (clusters). Du fait de la variabilité du nombre de points d'accès à portée, il est nécessaire de se restreindre à un sous-ensemble de points d'accès quand on détermine la distribution jointe pour chaque point de référence. Les auteurs du système HORUS proposent de ne considérer que les k points d'accès les plus puissants mesurés. Le nombre k est paramétré à l'usage du système HORUS. Les auteurs de HORUS forment l'hypothèse que les distributions des SS de chaque point d'accès sont indépendantes entre elles. Cela permet d'affirmer que la distribution jointe des k points d'accès est déterminée par l'ensemble des distributions séparées des points d'accès. L'intérêt de cette affirmation est de réduire la quantité de données nécessaires pour établir la distribution jointe des SS.

L'étape de regroupement (clustering) consiste à grouper les points de référence selon qu'ils ont des distributions jointes similaires. Les auteurs choisissent aussi de se limiter à la comparaison d'un nombre limité de points d'accès, ici q . Le nombre q est déterminé par paramétrage également. Pour affirmer que deux points de référence appartiennent à un même groupe, il est nécessaire que les q points d'accès pour lesquels la SS est la plus forte soient les mêmes pour les deux points de référence. L'affectation des points de référence à un groupe ne dépend pas de l'ordre des puissances reçues, uniquement de leur appartenance ou non aux q plus puissants.

La localisation fonctionne en trois temps. Premièrement, le terminal mobile mesure les SS des points d'accès à portée. Deuxièmement, grâce aux q points d'accès les plus puissants, le groupe correspondant est sélectionné. Troisièmement, parmi les points de référence du groupe sélectionné, la formule de Bayes est utilisée pour trouver lequel est le plus probable, connaissant la mesure du terminal mobile. La précision du joint clustering est de 2 mètres pour un intervalle de confiance de 80 %.

La seconde détermination de la position d'un terminal mobile est la triangulation incrémentale *Incremental Triangulation* [22]. Elle consiste à établir la distribution des puissances pour chaque couple point d'accès - point de référence. Ensuite, par ordre décroissant des puissances reçues depuis chaque point d'accès, on cherche tous les points dont la probabilité d'observer cette puissance dépasse un seuil fixé en paramètre. On élimine pour chaque point d'accès les points de référence qui ne sont pas assez probables. L'élimination de points de référence a lieu jusqu'à n'obtenir plus qu'un résultat, ou arriver au dernier point d'accès mesuré. Dans le dernier cas, on choisit alors le point de référence le plus probable parmi l'ensemble restant. Pour un

seuil de probabilité égal à 0,4, 4 points d'accès NAP=4 et un nombre maximum de points de référence retenus égal à 12, HORUS obtient une précision d'environ 1,8 m à 80 %. Considérer une distribution des puissances suivant une loi gaussienne [23] permet d'augmenter la précision d'environ 50 %.

Nibble [24] repose sur l'inférence bayésienne. Ce système de géolocalisation identifie la salle dans laquelle le terminal mobile est présent. Les probabilités sont calibrées par des mesures. Dans le cas présent, la puissance reçue est décomposée en 4 intervalles qui forment un ensemble discret de 4 valeurs. Le reste du système fonctionne comme un réseau bayésien classique. Les expérimentations de Nibble montrent une précision atteignant 97% dans la détermination de la présence du terminal mobile dans une salle. Cependant, les salles sont éloignées dans le bâtiment et le choix entre deux salles proches est difficile pour Nibble.

A. Haeberlen et al. proposent un système de géolocalisation qui utilise soit une distribution des puissances en histogrammes, soit selon une loi de Gauss [25]. Tout comme Nibble, la méthode utilisée fonctionne sur le principe d'un réseau bayésien. Les mesures sont cependant moins discrétisées que dans le système Nibble. La précision obtenue par l'utilisation d'une loi de Gauss est de 97%. La précision de la distribution en histogramme est de 95%. L'ajout d'un modèle de Markov (cf. section A.4) permet le tracking des terminaux mobiles. Le modèle de Markov indique les connexions entre les salles et détermine la position *a priori* dans le réseau bayésien. La construction du modèle de Markov se fait de la façon suivante : la probabilité de la transition d'un état vers lui-même est fixée manuellement. Les autres transitions reçoivent des probabilités identiques.

1.5.2 Utilisation d'un modèle de propagation

La seconde famille de techniques de géolocalisation s'appuie sur la modélisation de la propagation des ondes radio. La modélisation de la propagation et, plus particulièrement de l'atténuation, permet la détermination de la distance entre le terminal mobile dont on doit déterminer la position et différents points dont les coordonnées sont connues. La position peut ensuite être résolue analytiquement grâce à la trilatération.

L'article présentant SNAP-WPS (*Satellite Navigation and Positioning Group - Wireless Positioning System*) [26] démontre premièrement la possibilité d'utiliser la puissance du signal pour géolocaliser un appareil. Des mesures de puissance du signal réparties sur une durée de 24 heures à la fréquence de 2 mesures par seconde sont effectuées. Le résultat montre un signal plus stable aux heures où le bâtiment est vide alors que la puissance fluctue durant les heures de travail, en présence de personnes qui se déplacent dans les locaux. Cependant, l'écart type n'est que de 2,26 dBm, même en incluant ces heures de pointe. Cet écart type peu élevé prouve qu'il est possible d'exploiter la puissance du signal pour géolocaliser un terminal mobile.

Deuxièmement, la relation de Friis (cf. section A.1.3) est présentée. Elle permet de calculer la perte de puissance du signal dans un espace sans obstacle en fonction de la distance entre émetteur et récepteur et de la fréquence d'émission. Cependant, cette relation ne tient pas compte des obstacles qui existent entre la source et l'émetteur dans le cas de la propagation à l'intérieur d'un bâtiment. La suite de l'article considère que le lieu de test est un espace uniforme, donc idéal. Y. Wang fait une série de mesures à des points dont les coordonnées sont connues. Les points de l'échantillon de mesures sont placés dans un repère qui exprime la puissance de signal

reçue en fonction de la distance. Ces données sont ensuite utilisées pour déduire une relation mathématique entre la distance et la puissance du signal. La relation mathématique est obtenue par régression polynômiale. Dans le système SNAP-WPS, les polynômes de degré 1 à 6 ont été testés. Il apparaît que le troisième degré est le meilleur compromis entre précision et temps de calcul de la régression.

Une problématique secondaire de la géolocalisation basée sur Wi-Fi est soulevée dans ces travaux : quelle est la position optimale des points d'accès pour l'obtention d'une précision maximale ? L'auteur défend l'idée que, dans le cas de 3 points d'accès, plus le triangle formé par les trois points approche la forme d'un triangle équilatéral, plus la précision est grande. Il arrive à une précision de l'ordre des 3 mètres en alignant presque les points d'accès. En les plaçant de manière à former un triangle équilatéral, la précision parvient à 1,41 m. Une erreur de 1,41 m permet de localiser un terminal mobile dans un bureau.

Enfin l'étude de l'importance des murs et autres obstacles solides dans la propagation des ondes radios a été intégrée au modèle. Comme les auteurs de RADAR, Y. Wang utilise les travaux de Seidel et Rappaport [19]. La formulation diffère légèrement de celle employée dans RADAR :

$$P_d = p_{d_0} - 10n \log\left(\frac{d}{d_0}\right) + WEF$$

où :

- P_d est la puissance réelle reçue par le récepteur à une distance d de l'émetteur ;
- p_{d_0} est la puissance reçue à une distance de référence d_0 ;
- n est le taux de perte de puissance sur la distance ;
- d est la distance entre l'émetteur et le récepteur ;
- WEF est le *Wall Effect Factor*.

Pour déterminer le WEF, des mesures sont nécessaires. Deux points d'accès sont placés de chaque côté d'un mur, à la même distance de celui-ci. Puis un mobile mesure la puissance des signaux reçus pour chacun des deux points d'accès en se plaçant près du mur, d'un côté et de l'autre. La différence entre les deux mesures est induite par le mur uniquement car les distances sont les mêmes. Appliquées à un mur en brique, ces mesures quantifient le WEF à 15,9 dBm. Sur une cloison calcaire, cette mesure a pour résultat un WEF de 3,4 dBm.

La précision va de 1 à 3 mètres en pratique et est d'environ 0,1 m dans une situation idéale. Les auteurs envisagent dans des travaux futurs l'intégration au calcul de l'effet d'obstacles sur la puissance mesurée.

L'article [27], paru dans le domaine de la sécurité, a pour but premier de localiser les mobiles pirates qui tenteraient d'infiltrer un réseau informatique via sa partie sans fil. En effet, les réseaux Wi-Fi posent des problèmes de sécurité importants. Un point d'accès installé dans la zone sécurisée d'un réseau de type Ethernet filaire permet à des clients extérieurs de se connecter au réseau sans passer par le pare-feu. Ainsi, des clients non autorisés peuvent accéder au réseau d'une entreprise et pirater les données confidentielles qui y sont stockées.

Pour résoudre cette faille de sécurité, les auteurs de [27] se sont orientés vers l'établissement d'un modèle théorique de la propagation de l'onde radio comme présenté dans [26] qui transporte les données. Cependant leur approche est légèrement différente en ce sens qu'elle utilise directement la relation de Friis. La résolution des failles de 802.11 proposée dans l'article s'applique

aux points d'accès et aux cartes actifs sur le réseau. Dans le cas d'un terminal équipé d'une carte Wi-Fi qui écoute passivement les fréquences de la norme 802.11, il est impossible de détecter la position ou l'existence de celui-ci. La seule possibilité reste l'emploi de protocoles sécurisés tels que WPA (*Wi-Fi Protected Access*) défini pour Wi-Fi ou SSH (*Secure Shell*), protocole très connu et basé sur l'algorithme RSA.

Pour établir ce modèle, ils ont fait des relevés de puissance de signal à différents points de différents bâtiments. Ces mesures ont été effectuées grâce aux wireless tools [28], un outil libre dont les sources sont disponibles. Ces mesures leur ont permis d'établir un modèle de propagation des ondes se basant sur la relation de Friis mais s'appliquant aux milieux non vides. La relation de Friis ne s'applique que dans un espace sans obstacle (notre atmosphère par exemple).

La formule de Friis (cf. section A.1.3) peut aussi s'écrire sous forme logarithmique :

$$10 \log(P_R) - 10 \log(P_T) = 10 \log(G_R) + 10 \log(G_T) + 20 \log\left(\frac{\lambda}{4\pi}\right) - 20 \log(d)$$

où :

- P_R et P_T sont respectivement la puissance mesurée par le récepteur et la puissance d'émission ;
- G_R et G_T sont respectivement les gains d'antennes du récepteur et de l'émetteur ;
- λ est la longueur d'onde du signal ;
- et d est la distance entre l'émetteur et le récepteur.

Les auteurs considèrent le coefficient de 20, qui est multiplié par le logarithme de la distance. Ce coefficient est 2 pour le carré dans l'expression d'origine, multiplié par 10 car le travail en dBm fait apparaître un coefficient 10 pour le passage en "déci". Les auteurs de cet article ont aussi procédé à des régressions polynômiales sur les échantillons mesurés mais c'est la modification de Friis qu'ils ont préféré utiliser dans leur modèle.

Il leur apparaît alors que le coefficient 2 peut être changé pour diminuer la distance équivalente à une puissance du signal donnée car le milieu dans lequel l'onde se propage absorbe plus celle-ci. En effet, la puissance du signal est inversement proportionnelle au carré de la distance. Si l'on augmente la puissance au sens fonction mathématique appliquée à la distance, le signal s'affaiblit plus vite, comme il le fait réellement dès que l'on quitte les conditions d'application de la relation de Friis. Donc, la relation réciproque de Friis modifiée

$$d = 10^{\frac{10 \log(G_R) + 10 \log(G_T) + 20 \log(\frac{\lambda}{4\pi}) - 10 \log(P_R) + 10 \log(P_T)}{20}}$$

sera aussi en adéquation avec la réalité.

La précision obtenue est variable. La majorité des cas permet d'obtenir une précision de test de l'ordre de 1 à 3 mètres, mais certains cas particuliers donnent lieu à des erreurs supérieures à 8 mètres. Ceci est expliqué par les auteurs de l'article de plusieurs façons :

- les mesures dans le cas peu précis de l'Orinoco⁴ sont groupées à distance moyenne du point d'accès. Par conséquent, la précision sur de longues distances peut être mauvaise ;

⁴Orinoco est une marque de cartes Wi-Fi basées sur le chipset Lucent Wavelan très utilisé dans les mesures de puissance en vertu de sa précision.

- le point d'accès est dans une pièce contenant beaucoup d'éléments métalliques.

Devant ces points particuliers, la fin du rapport d'Interlink Networks (IN) préconise de continuer à étudier le modèle, en particulier par une vérification de la pertinence du coefficient de la distance dans la variante de la relation de Friis.

Pour les deux systèmes de géolocalisation, la position du dispositif à localiser est calculée par trilatération.

1.5.3 Réseaux de Neurones

Il est aussi intéressant de noter que des travaux [29] ont été faits sur la géolocalisation de mobiles équipés de cartes Wi-Fi en s'appuyant sur des réseaux de neurones pour l'exploitation des résultats. En se basant sur une approche de trilatération par des méthodes semblables à RADAR [15], un système a été proposé pour géolocaliser des mobiles Wi-Fi dans un WLAN (Wireless LAN, c'est à dire un réseau local sans fil) en mode infrastructure⁵. Les points d'accès constituent le point de communication entre le réseau hertzien et le réseau filaire. Les auteurs de ces travaux se sont basés sur l'architecture multi-couches de Perceptrons (MLP pour *Multi Layer Perceptron*) pour résoudre la position de façon plus fine et permettre l'apprentissage par ce système de la propagation des ondes pour affiner la résolution.

Pour cela, le modèle utilisé comporte une couche d'entrée qui donne les puissances reçues de différents points d'accès, au moins trois, une couche de sortie qui retourne les coordonnées $(x; y)$ dans le plan du mobile et une couche cachée de neurones (unité fonctionnelle) qui traite les données de la couche d'entrée pour transmettre ensuite le résultat à la couche de sortie. Pour établir le modèle, 194 mesures sont prises dans le bâtiment où ont lieu les tests et ces mesures sont fournies au réseau de neurones pour apprentissage initial. On peut alors fournir des mesures quelconques en entrée. Ces mesures sont alors traitées à l'aide du résultat de l'apprentissage pour déduire la position du mobile.

Ce modèle de géolocalisation est ensuite comparé à un modèle de *k plus proches voisins* pour déterminer sa précision. Pour rappel, l'algorithme des *k plus proches voisins* fonctionne par une recherche des *k* positions les plus proches de la mesure. La notion de proximité est définie selon des critères adaptés à l'objectif de l'algorithme. Par exemple, les *k* points choisis lors d'une itération par RADAR pour l'application de Viterbi sont les *k* positions les plus proches de la mesure où la proximité est déterminée par la distance euclidienne dans l'espace du signal. Une moyenne est appliquée sur les *k* résultats sélectionnés pour déterminer le résultat de l'algorithme.

Il s'avère alors que la précision des réseaux de neurones est proche de la meilleure précision possible avec les *k plus proches voisins*. Sachant que l'algorithme des *k plus proches voisins* est simple, cela montre que la modélisation de la propagation des ondes radio dans un bâtiment est un phénomène difficile à modéliser par une structure en réseau de neurones. L'approche basée sur des réseaux de neurones coûte cher en temps de calcul et n'apporte de plus qu'un gain mineur par rapport aux *k plus proches voisins*.

⁵Le mode infrastructure est le mode dans lequel les terminaux mobiles se rattachent à un point d'accès par lequel toutes les communications vont passer par opposition au mode ad-hoc dans lequel les terminaux communiquent deux à deux sans hiérarchie.

1.6 Analyse

Concernant la symbologie, celle-ci dépend des objectifs de la géolocalisation. Une symbologie fine doit être réservée à un ensemble d'applications qui requièrent de la précision. On distinguera deux types de symbologies : les symbologies sémantiques, qui portent un sens pour le système de géolocalisation et les applications qui l'utilisent, et les symbologies physiques, qui s'identifient à la géographie des lieux sans y associer le sens. Un exemple de symbologie sémantique est le découpage selon les pièces d'un bâtiment, chacune des pièces pouvant porter un sens, c'est à dire une application particulière. L'exemple trivial de symbologie physique est la représentation des positions par leurs coordonnées cartésiennes. La plupart des symbologies sémantiques peuvent s'établir en surcouche d'une symbolique physique.

Les caractéristiques des erreurs sont leur type et leur quantification. Les erreurs sont classifiables en erreurs aléatoires et en erreurs systématiques. Les erreurs aléatoires interviennent n'importe où et n'importe quand. On pourra souvent les traiter statistiquement. Les erreurs systématiques ont lieu à des endroits ou des temps précis et sont toujours les mêmes. La quantification des erreurs peut répondre à une loi de probabilité ou une valeur connue. La quantification est généralement calculée en mètres, mais il est aussi possible de l'exprimer (dans le cas de symbologies sémantiques notamment) par des pourcentages d'adéquation des réponses du système.

La vitesse de rafraichissement du système de géolocalisation dépend plus de la capacité de calcul de l'appareil qui calcule la position du terminal mobile. Elle est également dépendante de la distribution du système de géolocalisation. En effet, un système dont les calculs sont effectués sur un serveur centralisé s'expose à réduire la fréquence de rafraichissement quand le nombre de clients connectés augmente. À l'inverse, un système de géolocalisation distribué est plus réactif car chaque appareil ne supporte qu'une partie des calculs. Cependant, la fréquence de rafraichissement de la position d'un terminal mobile peut chuter si les communications liées à la géolocalisation sont trop importantes entre les appareils du réseau.

Nous prenons en compte deux coûts : le coût financier et le coût en temps. Le coût financier est dépendant du prix unitaire des équipements à installer pour que le système de géolocalisation fonctionne. Il est également dépendant du temps d'installation du système, car il est nécessaire de payer les personnes chargées de son installation. L'aspect financier est également présent dans la quantité d'opérations de maintenance et leur durée, du fait de la rémunération des personnes affectées à la maintenance. Les techniques basées sur Wi-Fi ont un coût financier peu élevé car elles utilisent une infrastructure réseau déjà présente. Les techniques de positionnement par satellite ont un coût peu important, chaque terminal devant être équipé d'un récepteur, et les satellites pouvant servir un nombre illimité de terminaux mobiles. Les techniques basées sur des appareils dédiés déployés dans l'enceinte des bâtiments coûtent cher car il est nécessaire d'acheter du matériel et de l'installer.

Le coût en temps est décomposé en deux éléments. Le premier élément concerne le temps de mise en œuvre du système de géolocalisation, par exemple la mise en place et la connexion des appareils nécessaires, ainsi que toutes les opérations nécessaires à l'initialisation du système de géolocalisation. Le second élément concerne les opérations de maintenance. On remarque que le coût financier est très corrélé au coût en temps. Les systèmes de géolocalisation par satellite sont les plus rapides à mettre en œuvre. Il suffit de connecter le récepteur pour se positionner.

L'extensibilité est principalement liée au coût d'installation. Active Bat coûte cher financière-

rement pour un déploiement global et RADAR coûte cher en temps dans le cas d'une extension de la zone de déploiement. Les systèmes centralisés coutent plus cher à étendre que les systèmes distribués. En effet, une compétition pour l'accès aux ressources critiques se met en place. Ce n'est pas ou peu le cas pour les systèmes distribués. Il s'agit de l'effet principal de la distribution du calcul à plusieurs acteurs du système. Les autres paramètres ne sont pas touchés par le choix de la centralisation ou du calcul distribué.

Système	Symb.	Erreur	Type	Coût	Ext.	Dyn.
RADAR	C	2 (moy.)	R	A	N	N
HORUS (IT)	C	1,8 (80%)	R	A	N	N
Ekahau	C	2,5 (moy.)	R	A	N	N
Fingerprinting	C	1,78 (moy.)	R	A	N	N
Nibble	S	40 à 100 %	R	A	N	N
SNAP-WPS	C	1 à 3	T	P	O	N
Interlink Networks	C	1 à 3	T	∅	O	N

FIG. 1.5 – Tableau récapitulatif des systèmes de géolocalisation Wi-Fi les plus importants.

Le tableau 1.5 indique pour les systèmes les plus importants quels sont leurs attributs. La colonne *Symb.* décrit la symbologie du système selon qu'il donne des coordonnées (C) ou la présence dans une salle (S). La colonne donnant l'erreur exprime cette dernière soit en mètres, soit en pourcentage de réussite pour les systèmes à symbologie de salle. La colonne *type* définit le type de modèle : basé soit sur le référencement d'une cartographie des puissances (C) soit sur la trilatération (T). La colonne *coût* prend pour valeur soit A pour l'acquisition des données, soit P pour une faible quantité de données requises, soit ∅ pour aucun coût de déploiement. Les colonnes *ext.* et *dyn.* expriment si le système est extensible (respectivement dynamique) ou pas (Oui ou Non). La dynamicité est la capacité du système à se corriger au fil du temps.

On remarque que les systèmes au sein d'une famille ont les mêmes caractéristiques dans les grandes lignes. En particulier, les modèles basés sur une cartographie des puissances affichent une meilleure précision que les modèles basés sur la trilatération. En revanche, ces derniers sont plus dynamiques car ils nécessitent moins de travail lors de leur mise en œuvre. Les systèmes les plus intéressants en ce qui concerne notre problématique sont :

- RADAR, dont les erreurs sont limitées ;
- Interlink Networks qui se déploie sans autre prérequis que les coordonnées des points d'accès ;
- SNAP-WPS, dont les erreurs sont moindre que pour Interlink Networks, mais qui requiert des données préalables afin de calibrer son expression de la distance en fonction de la puissance reçue.

Conclusion

Dans ce chapitre, nous étudions de nombreux systèmes de géolocalisation. Nous présentons tant des systèmes de géolocalisation basés sur Wi-Fi que des systèmes de géolocalisation basés sur d'autres supports. En effet, les méthodes employées dans chacun de ces systèmes peuvent

se rapprocher d'un support à l'autre. Elles permettent également de tenir compte des systèmes existants ainsi que de leurs inconvénients et avantages dans le cadre de la géolocalisation *indoor*.

On constate, à l'issue de l'analyse des travaux de l'état de l'art, que les méthodes de géolocalisation basées sur Wi-Fi ne sont pas les plus précises. En effet, les méthodes basées sur des émetteurs ultrasons sont largement plus précises. Cependant, le coût des systèmes basés sur les ultrasons est plus important car un grand nombre d'appareils supplémentaires est requis. De plus, la question se pose également de la disponibilité de ces appareils, ainsi que de leur placement dans un environnement quotidien. Inversement, les systèmes basés sur Wi-Fi sont certes moins précis, mais ils emploient une infrastructure pré-existante ou, du moins, utilisable à d'autres fins parallèlement à l'emploi pour la géolocalisation. De plus se pose la question de l'utilisation du système de géolocalisation, donc de la précision désirée. En effet, malgré une précision moindre que celle des systèmes basés sur les ultrasons, la précision des systèmes basés sur Wi-Fi peut s'avérer suffisante pour une majorité d'usages. En particulier dans notre cas, une précision de l'ordre de quelques mètres est suffisante pour des applications qui visent à localiser un individu dans une salle, ou par rapport à la zone de couverture d'un point d'accès. Cependant, des limites existent encore dans le cas d'une précision requise supérieure, comme dans le cas de la visite virtuelle de musée [30].

Dans le compte-rendu du workshop *IEEE HotMobile 2006*, la question "*Has Localization been Solved?*" est posée. Un élément de réponse apporté est le suivant : "*The general consensus was that the systems would converge. Hazas responded that indoor systems requiring special hardware are too expensive. Hightower added that, if two systems are required, they should be implemented on the same device so that users don't have to carry more than one device. Fox argued that special infrastructure isn't required because Wi-Fi signal strength information will be available in virtually any building*". Il est clair que l'avenir des systèmes de géolocalisation en intérieur passe par Wi-Fi du fait de son déploiement de plus en plus important. Ainsi, il est seulement nécessaire de déployer le logiciel, le matériel étant déjà présent.

Chapitre 2

Système de géolocalisation dans les réseaux Wi-Fi en intérieur

Introduction

Dans ce chapitre, nous présentons nos contributions dans le domaine de la géolocalisation en intérieur dans les réseaux Wi-Fi. Nous nous intéressons à la partie du serveur de localisation, c'est-à-dire à la façon de calculer la position du terminal mobile à partir des informations de puissances des signaux. Nous considérons acquise l'étape des mesures, qui relève de problématiques techniques (accès au pilote de la carte, possibilité de mesure).

Nous avons choisi Wi-Fi d'une part pour son accessibilité et d'autre part pour ses portée et débit importants. En effet, la portée en intérieur d'un périphérique Wi-Fi est de l'ordre de plusieurs dizaines de mètres, pouvant atteindre la centaine de mètres en extérieur. Le débit atteint quant à lui 54 Mbps, voire 300 Mbps pour la norme 802.11n. Ceci permet d'envisager des applications incluant des médias riches.

L'état de l'art révèle que les deux familles de systèmes de géolocalisation ont des défauts différents. D'une part, la grande quantité de données mesurées ou calculées dans le cadre des systèmes basés sur le référencement de points de mesures permet à ces systèmes d'avoir une précision intéressante mais les rend en contrepartie plus lents à déployer et à adapter à des changements d'environnement. D'autre part, les systèmes de géolocalisation basés sur la trilatération grâce aux distances obtenues par une relation entre la puissance du signal et la distance sont rapidement déployables et adaptables mais souffrent d'une précision inférieure à celle des systèmes basés sur le référencement de mesures.

La précision des systèmes de géolocalisation étudiés dans l'état de l'art a été quantifiée par leurs auteurs. Nous allons plus loin dans cette étude en observant la relation entre la densité des points de références utilisés par les modèles basés sur une cartographie des puissances et leur précision. Nous étudions également l'impact d'une connaissance basique de la topologie sur la précision d'un système de géolocalisation. L'objectif de ces deux problématiques est de parvenir à un système de géolocalisation précis en utilisant le minimum possible de données *a priori*. Nous souhaitons limiter la quantité de données requises afin de palier la nature éphémère des réseaux sans-fil. En effet, un point d'accès est facile à déplacer et le modèle doit être réactif face à ce type d'événements.

Nous souhaitons développer un système de géolocalisation possédant les propriétés suivantes :

- ▶ la *simplicité* à installer et mettre en œuvre ;
- ▶ la *dynamacité* , c'est-à-dire l'aptitude à rester opérationnel face à des changements d'environnement ;
- ▶ une *précision* de localisation de l'ordre de 2 mètres ;
- ▶ une fréquence de positionnement élevée, de l'ordre de la dizaine de calculs de positions par seconde ;
- ▶ l'*extensibilité* tant en termes de zone de couverture qu'en termes de nombre de terminaux mobiles localisés ;
- ▶ un coût temporel d'installation et de calibration/collecte des données faible : respectivement de l'ordre de quelques minutes à quelques heures pour un déploiement complet dans un bâtiment contenant une vingtaine de salles respectivement si l'on ne connaît pas la topologie ou si on la connaît.

La dynamacité et l'extensibilité sont, toutes choses égales par ailleurs, dépendantes du temps de mise en œuvre, ainsi que l'analyse des systèmes de géolocalisation Wi-Fi le montre (cf. section 1.6). Pour remplir les objectifs que nous nous sommes fixés, nous nous concentrons sur la conception d'un système de géolocalisation ayant une fréquence de positionnement élevée, une bonne précision et un temps de mise en œuvre court.

Nous donnons comme objectif une précision définie par une erreur de l'ordre de 3 mètres car elle nous permet de détecter la présence d'un terminal mobile dans une salle. Nous visons une fréquence de calcul de la position de 10 Hz, sur un PC standard⁶, de sorte que l'on puisse facilement déployer le système sur une vaste flotte de terminaux, soit en utilisant un serveur plus puissant, soit en distribuant les requêtes de positionnement sur plusieurs machines. Nous souhaitons également une installation rapide, pour laquelle peu de mesures préalables sont requises (une par salle) et pour laquelle une connaissance succincte de la topologie est suffisante.

Au sommaire de ce chapitre, nous posons premièrement la problématique liée à l'hétérogénéité de la topologie dans un bâtiment et son impact sur les ondes radio. Deuxièmement, nous définissons un modèle de propagation des ondes radio en intérieur, basé sur la calibration de la formule de Friis [31]. Troisièmement, nous exposons une amélioration de l'algorithme de type Viterbi présent dans RADAR, afin de prendre en compte la topologie des bâtiments. Quatrièmement, nous combinons nos deux premières approches pour obtenir un *modèle hybride*⁷. Le modèle hybride est plus rapide à mettre en œuvre que les approches par cartographie des puissances mais il est néanmoins plus précis que les approches basées uniquement sur la propagation du signal. Cinquièmement, nous décrivons les expérimentations menées sur les modèles que nous avons proposés. Nos modèles sont comparés à des modèles de l'état de l'art.

2.1 Caractéristiques de l'intérieur des bâtiments

Dans le reste de ce chapitre, nous employons la topologie au sens de la géométrie et de la répartition des obstacles principaux d'un bâtiment. Les obstacles principaux sont les éléments

⁶CPU à 2 GHz, 1 Go de RAM

⁷à différencier du sens usuel de l'hybridation entre GNSS et géolocalisation Wi-Fi en intérieur, par exemple.

non mobiles. En particulier, les murs et les éléments fonctionnels (conduites d'eau, matériel électrique) sont rarement modifiés dans un bâtiment donné. Interlink Networks [27] propose de remplacer l'exposant par 3,5 dans la formule de Friis pour modéliser le fait qu'une onde radio s'atténue plus vite dans un bâtiment à cause des obstacles. Cette solution suppose que l'intérieur d'un bâtiment soit homogène à une légère variation près. Cependant, des observations tendent à infirmer ce postulat.

Dans cette section, nous décrivons tout d'abord le bâtiment Numérica, dans lequel nous avons effectué des tests relatifs à la propagation des ondes radio et à la géolocalisation en intérieur. Nous y observons l'impact de la topologie du bâtiment sur les mesures des puissances des signaux reçus. Ensuite, nous définissons le terme d'hétérogénéité de la topologie, ainsi que la notion de zone homogène au sein d'un bâtiment.

2.1.1 Observations pratiques : le cas de Numérica

Numérica est le bâtiment hébergeant le LIFC sur le site de Montbéliard. Le LIFC occupe deux étages dans une aile de Numérica. Nous présentons dans un premier temps les plans de ces deux étages et décrivons leurs principales particularités. Dans un second temps, nous exposons les mesures préliminaires que nous y avons conduites et nous en donnons une signification liée à la topologie.

Caractéristiques remarquables des locaux de test

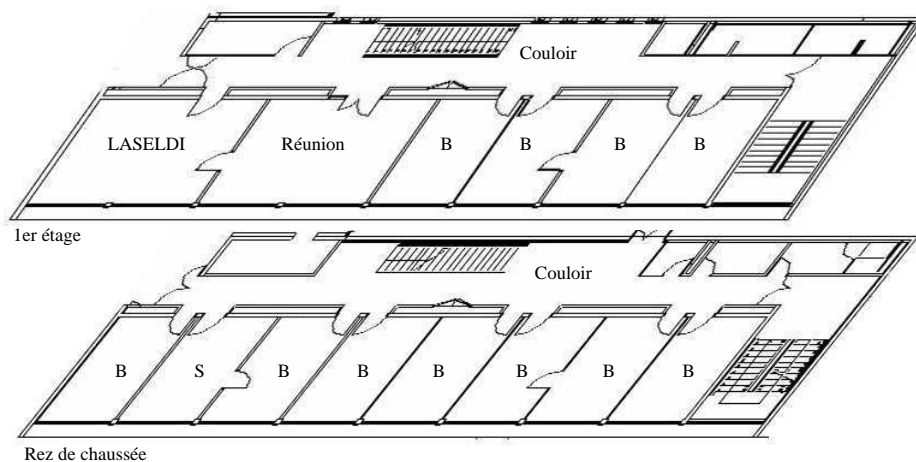


FIG. 2.1 – Plan 2D-isométrique des locaux du laboratoire LIFC, Numérica rez-de-chaussée et 1^{er} étage.

La figure 2.1 est un plan du rez-de-chaussée et du premier étage de Numérica, le bâtiment qui héberge les bureaux du LIFC à Montbéliard. On constate sur les deux étages que le bâtiment est séparé en deux dans la longueur par un double mur. Ce mur est un mur porteur épais, séparé par un vide d'une cloison d'épaisseur moindre. L'espace entre ces deux murs est occupé par des canalisations d'eau et des équipements électriques. D'un côté du mur central se trouve un couloir. Il ne comporte pas d'obstacles dans sa longueur et est de largeur variable. De l'autre

côté du mur central se trouvent les bureaux, indiqués par un B. Ils ont tous des dimensions comparables (environ $3 \times 4,5\text{m}$) et sont séparés par des cloisons. Certaines cloisons comportent une porte reliant deux bureaux. Au premier étage, deux pièces sont de dimensions différentes, environ le double d'un bureau. Ce sont la salle de réunion et le LASELDI (Laboratoire de Sémio-Linguistique, Didactique et Informatique). Des escaliers sont placés au centre du couloir et à l'extrémité du bâtiment.

Corrélation entre la topologie et la propagation des ondes

Les obstacles sont de nature variable, c'est-à-dire que les matériaux ou les dimensions diffèrent d'un obstacle à l'autre. Par conséquent, l'effet des obstacles sur la propagation des ondes est également variable. Les obstacles sont répartis irrégulièrement dans le bâtiment. Certaines salles (réunion et LASELDI) sont plus grandes que les autres. Ceci complique encore la modélisation précise du bâtiment. Aux obstacles variés s'ajoute un problème supplémentaire dans la modélisation du bâtiment : les meubles sont aussi des obstacles, et ils sont susceptibles d'être déplacés.

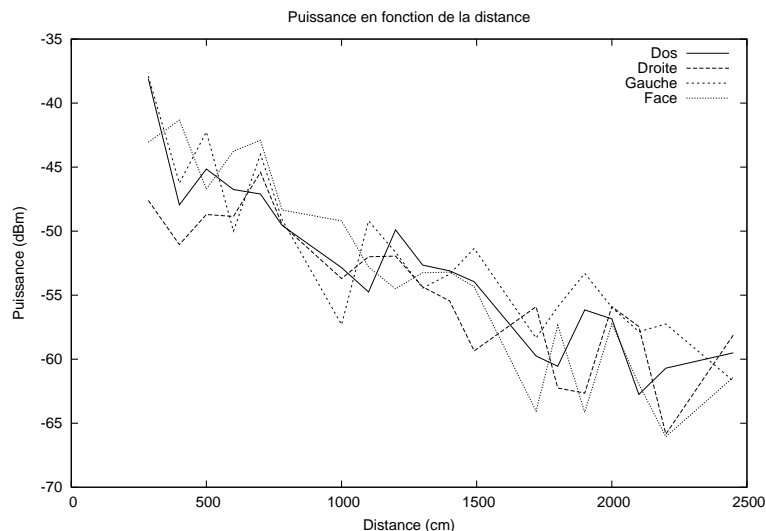


FIG. 2.2 – Atténuation du signal dans un couloir, puissance en fonction de la distance.

Les 4 courbes de la figure 2.2 illustrent l'atténuation du signal dans un couloir de Numérica. Chaque courbe correspond à une orientation du porteur de l'appareil Wi-Fi par rapport au point d'accès. Malgré l'absence d'obstacle entre le terminal mobile et le point d'accès, le signal ne se comporte pas comme on pourrait l'attendre en théorie, selon la formule de Friis. Globalement, la puissance diminue quand la distance augmente mais on observe plusieurs pics de puissance. Les pics sont plus importants quand la distance correspond à l'emplacement d'une armoire électrique ou de la plomberie dans le double mur central. On explique ce phénomène par l'effet des équipements sur la trajectoire des ondes, en particulier la réflexion du signal. Cette atténuation irrégulière rend difficile l'expression de la relation entre la puissance du signal reçu et la distance qui sépare le terminal mobile des points d'accès. La courbe de la figure 2.3 illustre ce point. Elle exprime la distance calculée par les modèles de Friis et d'Interlink Networks en

fonction de la distance réelle qui sépare l'émetteur et le récepteur. Dans cette figure, la courbe continue est la valeur réelle de la distance. Les deux autres courbes sont les distances calculées en chaque point avec la formule de Friis et la formule d'Interlink Networks. Les distances sont calculées en se basant sur les mesures de puissances des signaux exposées dans la courbe 2.2. On remarque que l'utilisation de Friis calcule une distance largement supérieure à la valeur réelle alors que le modèle d'Interlink Networks sous-évalue la valeur de la distance. Ces observations prouvent que dans un bâtiment, même en l'absence d'obstacles entre la source et le récepteur, le signal est fortement perturbé par l'ensemble de la topologie. On explique cela par l'ensemble des phénomènes qui influencent la propagation du signal radio : l'atténuation, la réflexion, la réfraction et la diffraction. Bien que l'échelle ne le laisse pas bien apparaître, l'erreur commise par le modèle d'Interlink Networks atteint facilement plusieurs mètres (cf. section 1.5.2).

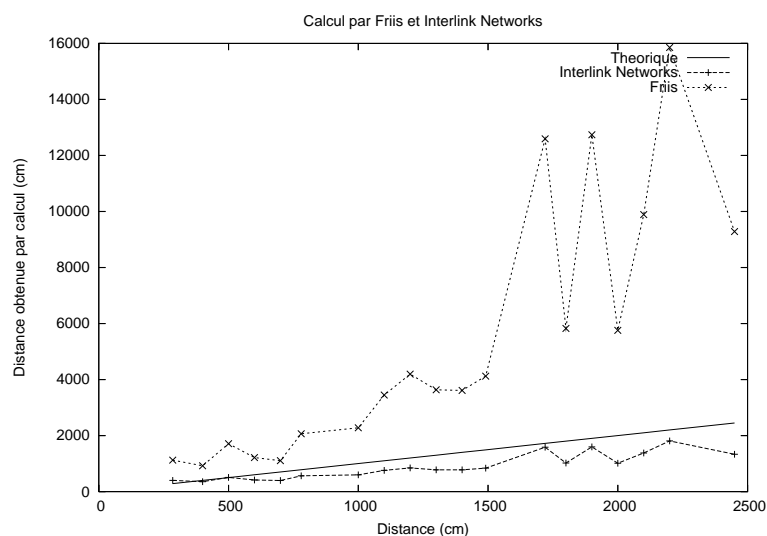


FIG. 2.3 – Comparaison du calcul de la distance : Friis et Interlink Networks.

Nous avons remarqué lors des mesures expérimentales que les personnes présentes dans le bâtiment forment un autre type d'obstacle car elles ont également un impact sur la propagation des ondes. En particulier, le porteur d'un appareil Wi-Fi est une source d'absorption du signal radio transmis. De plus, des événements ponctuels, organisés à Numérica et regroupant un nombre important de personnes, perturbent encore plus les mesures de puissance. Ceci a été aussi remarqué par Y. Wang [26] lors de mesures répétées de la puissance du signal.

Forts de ces observations et tenant compte de la propriété d'adaptabilité, nous avons décidé de ne pas faire reposer notre système de géolocalisation sur une modélisation précise de la topologie. En effet, le temps nécessaire à l'acquisition de toutes les données topologiques est trop important pour satisfaire le critère d'adaptabilité du système. De plus, l'environnement est sujet à modifications, en particulier du fait des variations du nombre de personnes présentes. C'est pourquoi nous nous orientons vers un modèle empirique qui pourra se corriger en cas de modification d'un ou de plusieurs paramètres : fréquentation, changement de disposition des meubles et déplacement d'un ou de plusieurs points d'accès.

2.1.2 Topologie hétérogène et découpage en zones homogènes

La topologie des bâtiments étant hétérogène, nous identifions deux degrés d'hétérogénéité. Nous définissons ensuite les zones homogènes et leur intérêt pour la géolocalisation en intérieur.

Hétérogénéité de la topologie

L'onde radio qui transporte le signal de l'émetteur au récepteur subit plusieurs phénomènes qui influent sur sa puissance. Le phénomène d'atténuation est lié à la distance parcourue par l'onde et se voit amplifié quand cette dernière traverse un obstacle. Les autres phénomènes ont lieu en présence d'obstacles sur la trajectoire des ondes. Ils modifient également la puissance reçue car ils altèrent la trajectoire de l'onde porteuse du signal. Pour cette raison, on parle de topologie hétérogène. En effet, selon la position du récepteur, le nombre et la nature des obstacles rencontrés par l'onde porteuse du signal varie.

Par conséquent, la mesure de la puissance d'un signal reçu ne s'exprime pas seulement par une fonction dépendante de la distance entre l'émetteur et le récepteur. Par ailleurs, comme deux points distincts placés à une même distance de la source du signal reçoivent un signal donné avec une puissance mesurée différente, la relation entre la puissance du signal et la distance n'est pas bijective. De ce fait, il est impossible de déterminer une expression exacte de la distance en fonction de la puissance du signal mesurée par le récepteur.

L'hétérogénéité de la topologie comporte deux niveaux :

- l'hétérogénéité intrinsèque à la composition des bâtiments ;
- l'hétérogénéité vis-à-vis des points d'accès.

L'**hétérogénéité** intrinsèque à la composition des bâtiments est liée à la répartition irrégulière des obstacles, tant en position qu'en nature, dans le bâtiment. En termes de nature, les murs sont composés de matériaux différents. Les matériaux influent sur la propagation des ondes radio. De la même façon, les équipements électriques et de la tuyauterie sont de tailles et de matériaux variables, influant également de manière variable sur la propagation des ondes radio. En termes de position, les murs ne sont pas placés à des distances identiques les uns par rapport aux autres. Cela influe également sur la propagation des ondes radio et rend la modélisation de la propagation ardue.

Zones homogènes

L'hétérogénéité de la topologie des bâtiments permet de définir des *zones homogènes* dans le bâtiment. Les **zones homogènes** sont des zones au sein desquelles la propagation des ondes radio est régulière. Nous choisissons de délimiter les zones homogènes par les murs présents. Entre les murs, c'est-à-dire dans une salle, l'indice de Friis est constant pour chaque point d'accès. Par conséquent, la régularité de la propagation de l'onde est définie par son atténuation suivant un unique modèle de Friis calibré (par point d'accès). La zone homogène telle que nous la définissons ne prend pas en considération la présence de personnes ou de meubles. En effet, malgré l'impact de ceux-ci sur la propagation des ondes, il est moins important que celui causé par les obstacles fixes. De plus, meubles et personnes peuvent se déplacer ou être déplacés et ne peuvent être modélisés.

L'hétérogénéité vis-à-vis des points d'accès concerne les zones homogènes. En effet, bien qu'au sein d'une zone homogène l'atténuation soit régulière, elle n'a pas été la même auparavant. Par conséquent les indices de Friis des points d'accès sont généralement différents pour une zone homogène donnée. On explique ce phénomène par les positions distinctes des points d'accès, dont les signaux rencontrent un nombre variable d'obstacles de natures variables. De fait, le modèle calibré appliqué pour chaque point d'accès est différent, c'est-à-dire que les atténuations ne sont pas homogènes entre les points d'accès. Nous parlons pour cette raison d'hétérogénéité vis-à-vis des points d'accès. La figure 2.4 décrit l'hétérogénéité vis-à-vis des points d'accès. Les traits pointillés représentent la trajectoire directe des ondes entre le point d'accès "UFC-DEA" et les points de calibration des bureaux. Les traits continus représentent la trajectoire directe entre le point d'accès "UFC-DEA" et les points de calibration situés dans le couloir. On remarque que les trajectoires vers les points de calibration des bureaux traversent un nombre de cloisons presque proportionnel à la distance entre l'AP et le point de calibration. Par opposition, les trajectoires de l'AP aux points de calibration du couloir interceptent sur une distance très différente le double mur central du bâtiment. Il ne faut pas oublier de plus que, plus l'onde touche le mur avec un angle important, plus elle va être réfléchi, ce qui atténue encore plus l'onde qui parvient au point de calibration.

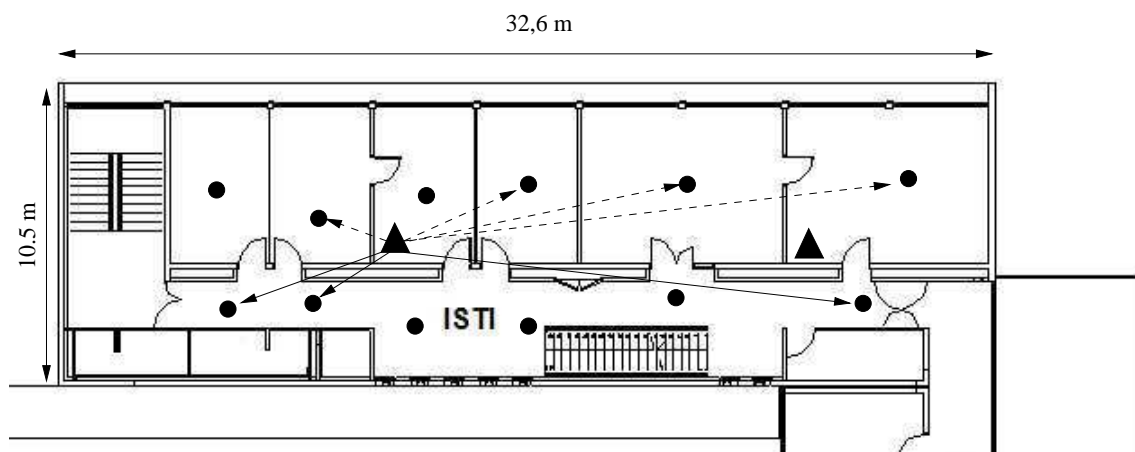


FIG. 2.4 – Disparité des trajectoires au sein des zones homogènes.

2.2 Calibration du modèle de propagation des ondes radio

Dans cette section, nous présentons le *modèle de Friis calibré*. Ce modèle est basé sur la variante de la relation de Friis présentée dans les travaux d'Interlink Networks [27]. Alors que ces travaux définissent une formule générique quel que soit le bâtiment, nous proposons de calibrer la formule de Friis en fonction de son lieu de déploiement. Nous présentons d'abord le modèle calibré de calcul de la distance en fonction de la puissance du signal mesuré (FBCM) [31]. Puis, nous décrivons un algorithme qui permet l'obtention d'une valeur approchée des coordonnées du terminal mobile, quel que soit le nombre de points d'accès disponibles. Nous tirons de ces deux points une conclusion sur les points forts et les limites du FBCM, du fait qu'il

est basé sur un modèle de propagation global au bâtiment.

2.2.1 Modèle de Friis calibré

Pour pallier l'hétérogénéité de la topologie, nous proposons un modèle basé sur la formule de Friis. Comme décrit dans les travaux d'Interlink Networks, nous remplaçons le carré de la distance par une puissance plus élevée. Cependant, contrairement à la formule proposée par Interlink Networks, nous proposons de calibrer la formule pour son usage dans un lieu précis. Premièrement, des mesures sont effectuées à des points dont les coordonnées sont connues. C'est la phase de calibration. Deuxièmement, les mesures obtenues permettent de calculer la formule exprimant la distance en fonction de la mesure de la puissance du signal.

La calibration requiert de connaître la position des points d'accès utilisés pour localiser les terminaux mobiles par le FBCM. La position est définie par les coordonnées du point d'accès dans le repère choisi pour exprimer les coordonnées des terminaux mobiles. Dans nos expérimentations à Numérica, nous nous basons sur un repère cartésien dont l'origine est l'angle en bas à droite sur le plan de la figure 2.1. Puis, les points nécessaires à la calibration, ainsi que leurs coordonnées, sont choisis. Pour chaque point de calibration, les valeurs de puissance des signaux reçus sont enregistrées, conjointement aux coordonnées du point. Comme on connaît également les coordonnées des points d'accès dont on a reçu les signaux, on peut calculer les distances euclidiennes de chaque point d'accès à chaque point de calibration. Avec N points d'accès utilisés et C points de calibration mesurés, $N \times C$ distances sont calculées. À chaque distance est associée la mesure qui lui correspond.

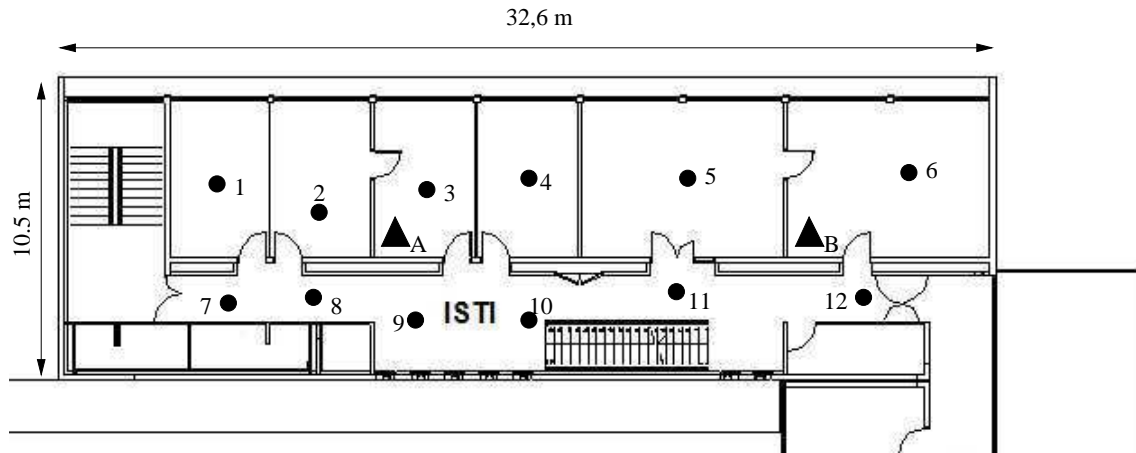


FIG. 2.5 – Positionnement des points de calibration et des points d'accès du banc d'essai n°1.

La figure 2.5 illustre le premier banc de test et le placement des points de calibration au premier étage du bâtiment Numérica lors de la première expérimentation. Les points sont les points de calibrations et les triangles sont les points d'accès. Le point d'accès A a pour ESSID "UFC-DEA" et le point d'accès B a pour ESSID "UFC-11G". Seuls figurent les points d'accès présents dans le bâtiment, à cet étage. Nous choisissons de placer un point de calibration par bureau ainsi que des points de calibration dans le couloir, environ à même hauteur que ceux des bureaux. Nous nous assurons ainsi de couvrir la plus grande variété de combinaisons d'obstacles

traversés, que ce soit en types (murs, etc.) ou en nombre. En assurant une large variété de scénarios de propagation, nous visons à obtenir un indice de Friis le plus représentatif de l'ensemble de l'étage. Grâce à un plan à l'échelle de Numérica et à la formule (2.2) dérivée de Friis, nous calculons les valeurs de l'indice de Friis.

L'expression de la distance en fonction de la puissance du signal est obtenue à partir de la relation de Friis, pour laquelle on remplace le carré de la distance par une inconnue. L'expression obtenue devient la suivante :

$$\frac{P_R}{P_T} = G_R G_T \left(\frac{\lambda}{4\pi} \right)^2 \frac{1}{d^i} \quad (2.1)$$

La puissance appliquée à la distance est i , que nous nommerons "indice de Friis". L'expression modifiée de Friis, une fois passée en échelle logarithmique, permet d'exprimer i par la relation suivante :

$$i = \frac{10 \log(G_R G_T) + 10 \log(P_T) + 20 \log\left(\frac{\lambda}{4\pi}\right) - 10 \log(P_R)}{10 \log(d)} \quad (2.2)$$

Le calcul de i est effectué pour chaque couple (point d'accès ; point de calibration) pour un ensemble de n mesures. L'indice de Friis final pour un point d'accès est égal à la moyenne des n indices de Friis obtenus. Du fait de l'hétérogénéité vis-à-vis des points d'accès, nous calculons un indice de Friis différent pour chaque point d'accès.

Point	SS	Indice	Distance (réelle)	Distance (IN)	Distance (FBCM)
7	-51,05	3,81	5,8	7,22	4,79
8	-50,4	5	3,7	6,92	4,64
9	-52,75	6,09	3,2	8,08	5,25
10	-54,15	4,06	6,2	8,85	5,65
11	-60,9	3,82	10,4	13,8	8,04
12	-67	3,67	16,8	20,62	11,05

TAB. 2.1 – Calcul d'indices de Friis et de distances, AP UFC-DEA.

Le tableau 2.1 contient le relevé des exemples de calcul de l'indice de Friis pour différents points du couloir du premier étage de Numérica. Il contient également les valeurs des distances calculées par l'indice unique et égal à 3,5 proposé par Interlink Networks, ainsi que les distances obtenues par le FBCM. À l'exception du point 12, l'évaluation de la distance par le FBCM est plus proche de la réalité que l'évaluation calculée par la formule proposée par Interlink Networks. On remarque que les valeurs de l'indice de Friis varient beaucoup d'un point de calibration à l'autre, allant de 3,67 à 6,09.

2.2.2 De l'analyse de l'atténuation du signal à la géolocalisation

À partir des distances séparant le terminal mobile de chacun des points d'accès, la position du terminal est calculée par trilatération. Comme nous l'avons exposé dans la partie des notions consacrée au calcul de la position d'un objet (cf. section A.2), la trilatération dans un espace de

dimension n se résout aisément avec $n + 1$ points connus et des distances exactes. Cependant, la distance calculée est en réalité différente de la distance qui sépare le terminal mobile de la source du signal. Par conséquent, les cercles de distance n'ont pas une intersection commune. Ils ont, deux à deux, zéro à deux points d'intersection. Dans ce paragraphe, nous exposons un algorithme qui calcule une valeur approchée de position du terminal mobile en se basant sur une grille de points dont on teste la distance par rapport aux cercles de distance.

Calcul du minimum de la distance maximale aux cercles

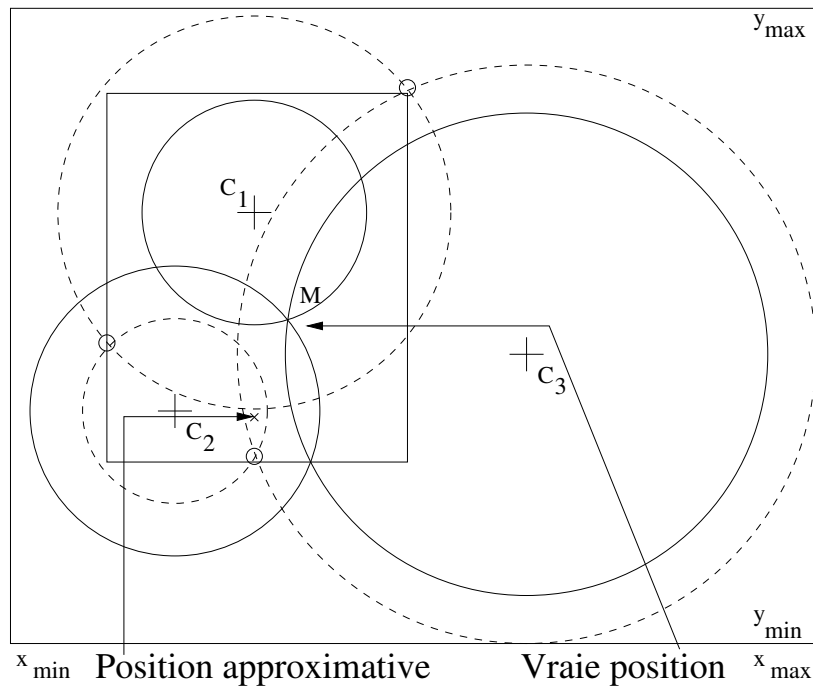


FIG. 2.6 – Trilatération avec des distances évaluées et biaisées.

Le FBCM ne permettant pas d'obtenir une distance exacte, les cercles de distance que nous utilisons ont pour rayon une approximation de la distance. Par conséquent, les cercles de distance n'ont pas une intersection unique commune à tous. De plus, si plus de $n+1$ cercles sont disponibles pour le calcul, il faut soit déterminer les coordonnées du point le plus proche par méthode analytique⁸, soit avoir une méthode de calcul approché. Nous nous sommes orienté vers la seconde possibilité.

Sur la figure 2.6, un exemple de calcul réel de la position d'un terminal mobile est donné. Les cercles continus représentent la distance réelle entre le terminal mobile et les points d'accès. Les cercles pointillés représentent les distances calculées entre le terminal mobile et les points d'accès. Avec les distances réelles, les 3 cercles ont un unique point d'intersection. Cependant, dans la réalité, ce sont les cercles pointillés qu'on observe. Par conséquent, la position du terminal mobile

⁸Une fonction de plusieurs variables permet de déterminer des dérivées partielles pour la recherche du point le plus proche. Cependant, la dérivée partielle dans le plan, avec 3 cercles, occupe plusieurs pages. C'est pourquoi nous avons choisi de simplifier le calcul par un algorithme de parcours du plan.

est approximée. On choisit le point du plan qui est le plus proche de l'ensemble des cercles. La figure présente un exemple de point le plus proche des 3 cercles. On peut également réduire l'espace de recherche du point le plus proche en se basant sur les points d'intersection les plus éloignés (ceux dont les coordonnées sont les plus extrêmes). Cette restriction est définie par le petit rectangle de la figure 2.6.

```

TRILATERATION(listeap, listedistances,  $x_{min}$ ,  $x_{max}$ ,  $y_{min}$ ,  $y_{max}$ , pas)
1  float min  $\leftarrow +\infty$ 
2  float xm  $\leftarrow x \leftarrow x_{min}$ 
3  float ym  $\leftarrow y \leftarrow y_{min}$ 
4  while  $x \leq x_{max}$ 
5  do  $y \leftarrow y_{min}$ 
6      while  $y \leq y_{max}$ 
7          do float  $d_{max} \leftarrow 0$ 
8              for int  $i \leftarrow 0$  to listeap.size()
9                  do if  $distance(x, y, listeap[i], listedistances[i]) \geq d_{max}$ 
10                     then  $d_{max} \leftarrow distance(x, y, listeap[i], listedistances[i])$ 
11                     if  $d_{max} \leq min$ 
12                         then  $min \leftarrow d_{max}$ 
13                          $xm \leftarrow x$ 
14                          $ym \leftarrow y$ 
15                      $y \leftarrow y + pas$ 
16                  $x \leftarrow x + pas$ 
17  return xm, ym

```

FIG. 2.7 – Algorithme de calcul du point le plus proche aux cercles.

L'algorithme *trilateration* est un algorithme de parcours du plan. L'idée directrice de l'algorithme est qu'un point qui est très proche de tous les cercles de distance est proche de l'intersection probable de tous les cercles s'ils avaient un unique point d'intersection (cf. section A.2.2). Cette notion de proximité avec l'ensemble des cercles de distance est traduite par la recherche du point dont la distance maximale par rapport à l'ensemble des cercles de distance est la plus faible. Nous lui affectons un pas p de parcours dans une zone à étudier. La zone à étudier est définie par les cercles de distance. L'algorithme parcourt une grille de points séparés par p mètres dans cette zone. Pour chaque point de la grille, l'algorithme calcule la distance entre lui et les périmètres des cercles de distance de chaque point d'accès. Il mémorise la valeur de distance la plus élevée, c'est-à-dire le cercle le plus éloigné du point. Le point sélectionné comme position du terminal mobile est le point pour lequel la distance mémorisée est la plus faible.

L'algorithme **trilateration**, décrit dans la figure 2.7, prend en paramètre la liste des points d'accès, la liste des distances, les bornes de parcours du plan, x_{min} , x_{max} , y_{min} et y_{max} , ainsi que le pas *pas* de déplacement dans le plan. La liste des distances, *listedistances*, contient les distances entre le terminal mobile dans l'ordre des points d'accès de la liste *listeap*. Un parcours du plan est effectué en testant les points d'une grille par leurs coordonnées $(x; y)$. Soit C un cercle de centre $O(x_O; y_O)$ et de rayon r . La distance d entre le point P de coordonnées $(x; y)$ et

le périmètre du cercle C est définie par $d = |\sqrt{(x - x_O)^2 + (y - y_O)^2} - r|$. La fonction *distance* de l'algorithme prend en paramètre les coordonnées du point courant, un point d'accès et la distance qui lui est associée et retourne la distance entre le point et le cercle de distance du point d'accès.

2.2.3 Limites

Compte tenu de l'hétérogénéité vis-à-vis des points d'accès, le FBCM souffre d'un défaut comparable, bien que moindre, à celui du modèle d'Interlink Networks. Bien qu'il adapte l'indice de la distance dans la relation de Friis, il prend pour hypothèse que l'intérieur du bâtiment pour lequel il a été calibré est assimilable à un environnement homogène. Cette hypothèse pourrait fonctionner avec des bâtiments dans lesquels les obstacles sont répartis régulièrement, c'est-à-dire qu'ils sont de même nature et que le nombre d'obstacles traversés est proportionnel à la distance entre l'émetteur et le récepteur.

Dans le cas de Numérica, cette affirmation se vérifie partiellement pour les points d'accès UFC-11G et UFC-DEA qui sont placés dans l'alignement des bureaux. Cependant, ce n'est pas le cas dans le couloir : en restant dans la même zone homogène, les signaux issus des points d'accès UFC-11G et UFC-DEA traversent un mur porteur, avec un angle variable qui provoque une réflexion variable et atténue différemment le signal. C'est ce qui explique que les indices de Friis calculés varient du simple au double selon les points. C'est également une source d'erreur dans la localisation puisqu'une mauvaise valeur de l'indice de Friis causera une mauvaise évaluation de distance. Par conséquent, le résultat de la trilatération comportera une erreur importante, de l'ordre de la dizaine de mètres dans les cas défavorables.

2.3 Prise en compte de la topologie dans les algorithmes de type Viterbi

Le FBCM étant insuffisant, nous proposons un modèle exploitant à la fois une base de données de points de référence et un modèle de propagation. Dans un premier temps, nous proposons d'améliorer l'algorithme Viterbi-like (cf. RADAR dans la section 1.5.1). En effet, l'algorithme, tel qu'il est présenté par V. Bahl, ne tient pas compte de la topologie. En effet, les distances considérées sont des distances euclidiennes. Or, il est fréquent que le déplacement du terminal mobile entre deux points l'oblige à emprunter des détours car il ne peut traverser les obstacles. Tenir compte de la topologie permet de mieux évaluer la pertinence du choix du point élu par l'algorithme Viterbi-like.

Dans cette section, nous proposons dans un premier temps une solution pour la prise en compte de la topologie par l'algorithme de Viterbi. Nous nous basons pour cela sur une modélisation de la topologie entre les points de la cartographie des puissances grâce à un graphe des connexions directes possibles entre points. Dans un second temps, nous proposons une méthode de calcul incrémentale de l'algorithme Viterbi-like pour en réduire la complexité. En effet, cette dernière est élevée lorsque les paramètres de l'algorithme, k et n , croissent. La combinaison des deux propositions forme l'algorithme *Fast-viterbi-like-improved* (FVLI). L'optimisation du calcul de l'algorithme est démontrée et sa complexité calculée.

2.3.1 Représentation de la topologie

Nous choisissons de représenter dans l'algorithme Viterbi-like la topologie grâce à un graphe de connexité des points de la cartographie des puissances. Dans le graphe de connexité, les sommets sont les points de référence. Pour deux sommets, s'il existe un arc alors il est possible de tracer un segment de droite entre eux sans intersection avec un élément fixe de la topologie. Les arcs sont étiquetés par la distance euclidienne entre les sommets qu'ils connectent.

La partie gauche de la figure 2.8 illustre la distance euclidienne entre les points A et B (ligne pointillée), notée AB , et la distance que parcourera véritablement le terminal mobile pour aller du point A au point B (ligne continue fléchée). Les distances euclidiennes sont faciles à déterminer à partir des coordonnées cartésiennes des points, l'inverse de la distance à parcourir pour un terminal mobile. La partie centrale de la figure 2.8 illustre le calcul de la distance entre les points en tenant compte de la topologie. Cette distance est nommée *distance topologique*. Lorsque deux points sont joints par un segment de droite sans intersection avec un obstacle, leur distance topologique est égale à la distance euclidienne. On dit que ces points sont *voisins*. Les points B et C sont dans ce cas. Les distances topologiques sont obtenues par une somme de distances connues. Par exemple, la distance topologique entre les points C et E est égale à la somme des distances CD et DE .

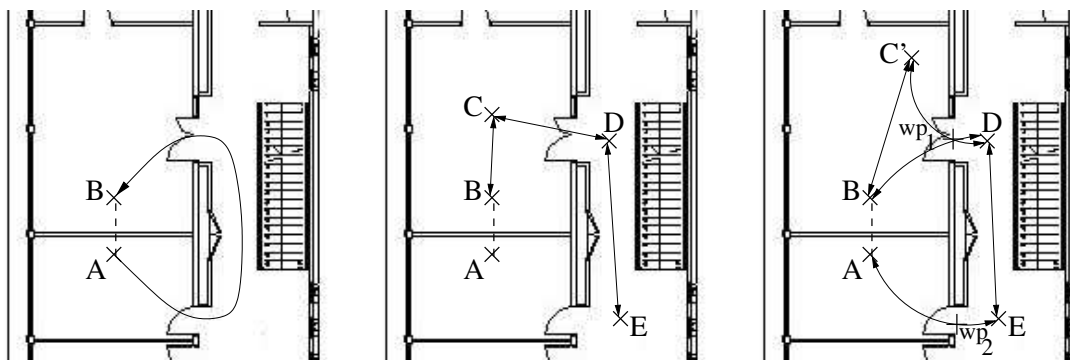


FIG. 2.8 – Représentation du calcul des distances en fonction de la topologie.

L'algorithme qui calcule les distances fonctionne en deux étapes :

- le calcul des distances euclidiennes entre les points **voisins** ;
- la somme des distances entre des points voisins afin de former un chemin de voisin en voisin entre deux points non voisins.

La première étape permet de former un graphe non orienté de points voisins. Dans ce graphe, les sommets sont les points de référence et les arêtes sont les possibilités de déplacement entre points voisins. Dans un bâtiment, si toutes les salles sont accessibles, le graphe des voisinages est fortement connexe. On représente les connexions entre les points voisins par une matrice M dont les coefficients $m_{i,j}$ sont les distances entre les points. On considère p points de référence. La première étape initialise la matrice :

- les coefficients $m_{i,i}, 0 \leq i < p$ sont égaux à O , la distance entre un point et lui-même étant nulle ;


```

DISTANCES_1( $M, \Omega(P)$ )
1  boolean Calculfait  $\leftarrow$  true
2  while Calculfait = true
3  do Calculfait  $\leftarrow$  false
4      for int  $i \leftarrow 0$  to  $\Omega(P) - 1$ 
5      do for int  $j \leftarrow i + 1$  to  $\Omega(P) - 1$ 
6          do if  $(i \neq j)$  and  $(M[i, j] > 0)$ 
7              then for int  $k \leftarrow j + 1$  to  $\Omega(P) - 1$ 
8                  do if  $(M[j, k] > 0)$  and  $((M[j, k] + M[i, j] < M[i, k])$  or  $(M[i, k] = -1))$ 
9                      then  $M[i, k] \leftarrow M[j, k] + M[i, j]$ 
10                     Calculfait  $\leftarrow$  true

```

FIG. 2.9 – Calcul des distances topologiques minimales entre les points de référence.

- les coefficients $m_{i,j}, 0 \leq i < p, 0 \leq j < p$ tels que les points P_i et P_j sont voisins sont égaux à la distance euclidienne entre les points P_i et P_j ;
- le reste des coefficients de la matrice est égal à $+\infty$.

Le matrice M est symétrique car la distance entre deux points est la même quel que soit le sens de déplacement.

L'algorithme *Distance_1* de la figure 2.9 décrit le calcul des distances entre points non-voisins. L'algorithme fonctionne en calculant la distance entre un point et les voisins de ses voisins, puis avec les voisins suivants et ainsi de suite jusqu'à avoir atteint les points les plus lointains. Il peut y avoir plusieurs chemins entre deux points, c'est pourquoi on conserve le plus court. Dans le détail, l'algorithme parcourt la moitié supérieure de la matrice M . Pour chaque élément $m_{i,j}$, il se déplace à la i^e ligne de la matrice M . La i^e ligne de la matrice M contient en effet les distances connues entre le point d'indice j et les autres points. La lecture de la ligne commence à $j + 1$ (partie supérieure de la matrice M). Pour chaque coefficient différent de $+\infty$, on compare la somme $S = m_{i,j} + m_{j,k}$ avec la valeur de $m_{i,k}$. Si $S < m_{i,k}$ alors on a trouvé un chemin plus court entre les points d'indice i et k : $m_{j,k}$ prend S pour valeur. Dans le cas contraire, le coefficient $m_{i,k}$ reste inchangé. Le parcours de la matrice est effectué tant qu'au moins un remplacement de coefficient a été fait. Il faut au plus $\Omega(P)$ parcours de la matrice pour finir le calcul. Lorsque la moitié supérieure de la matrice M est calculée, sa moitié inférieure est obtenue en recopiant les coefficients symétriquement par rapport à la diagonale, c'est-à-dire $m_{i,j} = m_{j,i}, 1 \leq i < \Omega(P), 0 \leq j < i$.

Cet algorithme est suffisant lorsque les points de référence sont en visibilité les uns par rapport aux autres. Cependant, ce n'est pas toujours le cas. Dans le cas où tout ou partie des points de référence ne se "voient" pas entre eux et ne permettent pas de créer un graphe fortement connexe par extension, un ajout au graphe doit être effectué. L'ajout consiste en un ensemble L de nouveaux points. Chacun des points de l'ensemble relie deux points de référence. On détermine alors les voisinages entre les points de l'ensemble $P \cup L$. Le calcul des distances est le même que celui de l'algorithme 2.9, en travaillant sur l'ensemble $P \cup L$ au lieu de P . Quand la matrice des distances est complète, seules les lignes et colonnes correspondant aux points de référence sont conservées.

2.3.2 Réduction de la complexité de l'algorithme Viterbi-like

La complexité de l'algorithme Viterbi-like étant exponentielle, l'augmentation des paramètres k et n nuit à la réactivité de la localisation. Nous proposons une méthode optimale de résolution de l'algorithme Viterbi-like. D'abord, nous décrivons le fonctionnement de la méthode optimale qui met à jour progressivement les distances entre les ensembles de points. Ensuite, nous démontrons que le résultat de l'algorithme est la valeur exacte de la distance la plus courte. Enfin, nous comparons les complexités de l'algorithme Viterbi-like et de son optimisation.

Algorithme

Par la suite, on note $d(A, B)$ la distance entre deux points A et B . La distance est abordée dans cette partie au sens large, c'est-à-dire qu'elle peut être euclidienne, topologique ou toute autre distance selon la problématique à laquelle l'algorithme Viterbi-like est appliqué.

L'objectif de l'optimisation est de construire itérativement des sous-solutions optimales exhaustives à partir de la première paire d'ensembles de points. On considère l'algorithme Viterbi-like de paramètres K et N . À chaque itération, les K points sélectionnés sont les K points les plus proches d'une mesure dans l'espace des puissances. À la première itération, on constitue l'ensemble E_0 de K points $p_{0,i}$ avec $0 \leq i < K$. On constitue de même l'ensemble E_1 de K points $p_{1,i}$ avec $0 \leq i < K$. Pour chaque point $p_{1,i}$, on calcule les distances par rapport à tous les points de l'ensemble E_0 . La distance la plus faible trouvée est conservée dans un vecteur V_1 de K nombres réels. La distance minimale ayant pour destination le point $p_{1,i}$ est l'élément $V_{1,i}$ du vecteur des distances. À chaque itération suivante, on met à jour le vecteur des distances par les points du nouvel ensemble calculé. Soit V_i tel que $0 \leq i < N$ le vecteur des distances minimales des chemins arrivant aux points de l'ensemble E_i . Possédant V_i , E_i et E_{i+1} à l'itération $i + 1$, on calcule chaque élément de V_{i+1} de la manière qui suit :

$$\forall j \in [0, K[, V_{i+1,j} = \min(V_{i,l} + d(p_{i,l}, p_{i+1,j}))$$

avec $l \in [0; K[$.

Propriété

Chaque élément $V_{i,j}$ est égal à la distance du plus court chemin débutant à un point de E_0 et passant par un point de chaque ensemble E_i , $1 \leq i < j$ intermédiaire (équation 2.3).

$$V_{i,j} = \min_{\{p_{x,k_x}\}_{x=0,\dots,n-1}} (\sum d(p_{x,k_x}, p_{x+1,k_{x+1}})) / k_i = j \quad (2.3)$$

Démonstration

Soit P_i la proposition $V_{i,j}$ est la distance du plus court chemin passant par un point de chaque ensemble de l'historique et se terminant au $j^{\text{ième}}$ point de l'ensemble d'arrivée. Soit $p_{0,k_{\min}}$, $k_{\min} \in [0; K[$, le point tel que, $\forall (j; k) \in [0; K]^2$, $d(p_{0,k_{\min}}, p_{1,j}) \leq d(p_{0,k}, p_{1,j})$. Donc, la proposition initiale $P_1 : V_1$ est le vecteur des distances minimales des chemins entre les points de l'ensemble E_0 et chaque point de l'ensemble E_1 , c'est-à-dire $V_{1,i} = \min(d(p_{0,j}, p_{1,i}))$ est vraie.

On considère que l'hypothèse P_{i-1} est vraie. La proposition P_i est-elle vraie ? Cette interrogation revient à se poser la question de savoir si

$$\min(V_{i-1,m} + d(p_{i-1,m}, p_{i,j})) = \min_{\{p_{x,k_x}\}_{x=0,\dots,n-1}} (\sum d(p_{x,k_x}, p_{x+1,k_{x+1}})) / k_i = j$$

Or,

$$\begin{aligned} & \min(d(p_{0,k_0}, p_{1,k_1}) + d(p_{1,k_1}, p_{2,k_2}) + \dots + d(p_{i-1,k_{i-1}}, p_{i,j})) \\ &= \min(\min(d(p_{0,k_0}, p_{1,k_1}) + d(p_{1,k_1}, p_{2,k_2}) + \dots + d(p_{i-2,k_{i-2}}, p_{i-1,m})) + d(p_{i-1,m}, p_{i,j})) \\ &= \min(V_{i-1,m} + d(p_{i-1,m}, p_{i,j})) \end{aligned}$$

Donc la proposition P_i est vraie.

Comparaison des complexités : FVLI et Viterbi-like

Nous comparons la complexité de l'algorithme FVLI à celle de Viterbi-like [16]. Pour cela, nous considérons que l'ensemble des opérations de FVLI est multiplié par le paramètre N de l'algorithme. En effet, il faut tenir un historique de toutes les exécutions en cours de l'algorithme Viterbi-like-improved du fait de l'aggrégation des distances dans les vecteurs V_i , $1 \leq i \leq N$. À chaque itération, nous calculons les distances entre deux ensembles E_{i-1} et E_i de K points, soit le calcul de K^2 distances. Chaque distance est sommée à la distance du chemin le plus court permettant d'arriver à l'élément de E_{i-1} , source de l'arc courant entre E_{i-1} et E_i . On a donc $K^2 \times N$ sommes de nombres réels. Les distances sont obtenues par l'accès à la matrice des distances entre les points de la cartographie des puissances.

N	K	Complexité de Viterbi-like (nombre de sommes)	Complexité de Viterbi-like improved (nombre de sommes)
3	3	54	27
4	4	768	64
5	5	12500	100
7	7	4941258	343
10	10	9.10^{10}	900

TAB. 2.2 – Comparaison des complexités des algorithmes Viterbi-like et Viterbi-like-improved.

Le tableau 2.2 compare les complexités de Viterbi-like et Viterbi-like-improved pour diverses combinaisons allant jusqu'à $K = 10$ et $N = 10$. Pour cette dernière, on remarque que le nombre d'opérations requises par Viterbi-like est divisé par 10^8 par l'optimisation de Viterbi-like-improved. L'importance du gain de vitesse prend son sens dans notre contexte de géolocalisation dans le cadre de la diffusion de contenus multimédia, requérant des traitements temps-réel. De plus, plusieurs clients doivent être supportés par un tel système, rendant encore plus critique la réduction des temps de calcul.

2.4 Fusion de la cartographie des puissances et du FBCM

Le calcul des indices de Friis du FBCM est très variable en fonction de la position du point de calibration dans le bâtiment, comme nous le montre le tableau 2.1. Par ailleurs, la géolocalisation basée sur une cartographie des puissances requiert beaucoup de données (cf. section 1.5.1). Nous proposons de combiner les deux approches dans un modèle hybride. Le modèle hybride permet d'utiliser une cartographie des puissances peu dense, de l'ordre d'un point de référence par salle. Le point de référence sélectionné est ensuite utilisé comme point de calibration unique dans le FBCM. Nous visons ainsi à utiliser une calibration plus adaptée et précise.

Le reste de cette section propose dans un premier temps une solution de collecte des données de puissance des signaux pour établir la cartographie des puissances. Dans un second temps, nous exposons l'idée directrice du modèle hybride basé sur le référencement et le FBCM (*FBCM and Reference-Based Hybrid Model* FRBHM) et justifions l'intérêt des deux variantes. Dans la troisième sous-section, nous définissons les trois variantes du FRBHM. La quatrième sous-section expose la modélisation de la topologie des bâtiments afin de rendre possible l'usage de l'algorithme FVLI dans le FRBHM continu.

2.4.1 Mesures de référence

Il est nécessaire de constituer une base de données de points de référence, c'est-à-dire la cartographie des puissances. La collecte des données s'effectue par des mesures des puissances des signaux reçus. Nous privilégions des mesures par rapport à une simulation par calcul car les algorithmes pour calculer la propagation des ondes radio dans un bâtiment sont, soit coûteux en temps de calcul, soit trop imprécis [15]. Un point de référence est défini par :

- ses coordonnées $(x; y; z)$ dans notre repère cartésien ;
- la liste des points d'accès captés, avec pour chacun la puissance du signal reçu. Cette donnée est l'ensemble $\{(AP_1; SS_{ref, AP_1}); (AP_2; SS_{ref, AP_2}); \dots; (AP_N; SS_{ref, AP_N})\}$;
- l'orientation o du terminal mobile lors de la mesure.

On observe des variations dans les mesures à un point donné. Afin de minimiser l'impact des variations, nous avons choisi de procéder à plusieurs mesures de puissance pour chaque couple (point d'accès ; point de référence). La valeur utilisée actuellement dans nos algorithmes est la moyenne des mesures. La moyenne est effectuée en convertissant les mesures en milliwatts. La moyenne des mesures valeurs en milliwatts est calculée puis reconvertie en dBm. La formule de la moyenne de m mesures ss_i en dBm est donc la suivante :

$$P_{moyenne} = 10 \log\left(\frac{\sum_{i=1}^m 10^{0,1 \cdot ss_i}}{m}\right) \quad (2.4)$$

2.4.2 Fonctionnement du FRBHM

D'une part, le FBCM est basé sur la moyenne d'indices de Friis. Nous avons remarqué que les indices de Friis varient jusqu'à un facteur 2 parmi l'ensemble des points de calibration (cf. tableau 2.1). Une telle variation conduit à une erreur de calcul d'une distance d atteignant en moyenne $d \cdot \sqrt{2} - d$. Par exemple, une distance réelle de 10 mètres sera évaluée à 30 mètres.

Comparée à la taille d'une salle dans un bâtiment, une erreur aussi importante ne permet pas de localiser un terminal mobile avec précision et ne permet donc pas de lui fournir des services contextuels.

D'autre part, les systèmes de géolocalisation nécessitent de mesurer les puissances pour un grand nombre de points de référence. En effet, les travaux basés sur la cartographie des puissances étudiés, se basent sur une grille de points de référence tous les mètres. Cela conduit à une prise de mesures proportionnelle à la surface du bâtiment.

Nous proposons de limiter le nombre de points de référence à 1 par salle. Dans des bureaux, ceci donne un point environ tous les 5 mètres. L'erreur ne peut être inférieure à $\frac{5\sqrt{2}}{2}$ mètres. Une telle erreur est encore inexploitable par des services contextuels. Le FBCM intervient à ce moment. Il est calibré par rapport à l'ensemble des mesures du point de référence sélectionné. Puis, la position du terminal mobile est calculée par trilatération. De cette façon, la calibration ne considère pas la moyenne des indices de Friis de tous les points de référence. Au contraire, elle est limitée à la zone homogène sélectionnée et elle est plus appropriée pour calculer des distances.

2.4.3 Algorithmes FRBHM

Dans cette sous-section, nous détaillons les trois variantes de l'algorithme FRBHM : le *FRBHM basique*, le *FRBHM avec prise en compte de la topologie* et le *FRBHM continu*. Le FRBHM basique combine simplement la cartographie des puissances et le FBCM. Le FRBHM avec prise en compte de la topologie intègre l'algorithme FVLI avec la cartographie des puissances puis déduit la position à l'aide du FBCM. Le FRBHM continu sélectionne un ensemble de k points applique le FBCM à chacun d'entre eux avant que l'algorithme FVLI ne sélectionne le point final.

Les trois formes d'algorithmes FRBHM fonctionnent sur une demande de localisation comportant une liste de points d'accès et des paquets associés dont les puissances ont été mesurées. Une mesure M est de la forme $M = \{(ap_1; ss_1); (ap_2; ss_2); \dots; (ap_N; ss_N)\}$. Dans chaque couple $(ap_i; ss_i)$, ap_i est l'adresse physique d'un point d'accès et ss_i est la mesure en dBm de la puissance du signal associé au point d'accès.

FRBHM basique

À la réception d'une demande de localisation, le FRBHM basique procède en deux étapes :

- le choix d'un point de la cartographie des puissances ;
- la calibration et l'application du FBCM pour calculer les coordonnées du terminal mobile.

L'algorithme 2.10 expose les étapes du FRBHM basique. Le FRBHM basique prend en paramètre la mesure envoyée dans la requête de localisation. Le nombre de points d'accès, nb_ap est déduit de la mesure et un tableau de nombres réels est créé pour contenir les indices de Friis. La fonction *ss_map* prend la mesure en paramètre et 1 pour spécifier qu'elle ne renvoie qu'un point de référence (le plus proche dans l'espace du signal). Le point renvoyé par la fonction est p . Par rapport au point p , les indices de Friis par rapport à chacun des points d'accès sont calculés par la fonction *friis_idx* et insérés dans le tableau *list_idx*. Le point sélectionné, appartenant à l'ensemble fini et dénombrable des points de référence est recalculé par trilatération (fonction

```

FRBHM(measure)
1  int nb_ap ← measure.size()
2  Array of nb_ap float list_idx
3  Point p ← ss_map(measure, 1)
4  for int i ← 1 to nb_ap
5  do list_idx[i] ← friis_idx(p, measure[i])
6  return FBCM(list_idx, measure)

```

FIG. 2.10 – Algorithme FRBHM basique.

FBCM. Le point résultant appartient à l'espace cartésien, un ensemble infini indénombrable. C'est ce point qui est retourné par l'algorithme FRBHM basique.

FRBHM tenant compte de la topologie

À chaque requête de localisation d'un terminal mobile, le FRBHM basique procède en trois étapes :

- dans la base de données des points de référence, il cherche les K points les plus proches de la mesure reçue. La proximité est calculée par tri croissant des distances euclidiennes dans l'espace du signal entre les points de référence et la mesure. Les K premiers points de la liste triée sont sélectionnés ;
- en utilisant les ensembles de K points de référence constitués lors des appels précédents au FRBHM, l'algorithme FVLI est exécuté et élit un point du dernier ensemble constitué. La précision de ce point dépend de la densité de la cartographie des puissances ;
- le FBCM est appliqué avec des indices de Friis calculés à partir des mesures du point de référence choisi. Il calcule des distances en se basant sur les mesures transmises lors de la requête de positionnement.

```

FRBHM(measure, fvli_list)
1  int nb_ap ← measure.size()
2  Array of nb_ap float list_idx
3  Array of kPoints k_pts_list ← ss_map(measure, 1)
4  queue(fvli_list, k_pts_list)
5  Point p ← fvli(fvli_list)
6  for int i ← 1 to nb_ap
7  do list_idx[i] ← friis_idx(p, measure[i])
8  return FBCM(list_idx, measure)

```

FIG. 2.11 – Algorithme FRBHM avec prise en compte de la topologie.

L'algorithme 2.11 explicite ces trois étapes. Il reçoit en paramètre la liste des mesures et l'historique requis par l'algorithme FVLI. L'algorithme stocke le nombre de points d'accès pour

lesquels on a une mesure de puissance lors de la requête de localisation. Puis, il constitue la liste des k points de référence choisis à partir de la mesure : la fonction *ss_map* prend en paramètre la mesure. Elle exécute une recherche des k plus proches dans l'espace des puissances (kNSS). La fonction *queue* prend en paramètres les k points choisis et les ajoute à l'historique des positions de l'algorithme FVLI. La fonction *foli* exécute l'algorithme FVLI sur les n ensembles de k points qui lui sont passés en paramètre. Elle renvoie le point choisi p , selon le fonctionnement du FVLI. La fonction *friis* prend en paramètre le point p et la mesure. Pour chaque point d'accès présent dans la mesure, elle calcule l'indice de Friis basé sur les mesures de référence de p . Enfin, la fonction *FBCM* calcule un nouveau point basé sur la liste des indices de Friis *list_idx*. Le point calculé est la valeur de retour de la fonction FRBHM.

Le passage du discret (cartographie des puissances) au continu (FBCM) n'intervient qu'à la fin de l'algorithme. Nous proposons d'améliorer le FRBHM basique pour que les points utilisés par l'algorithme FVLI soient déjà choisis dans l'espace cartésien. En effet, avec peu de points de référence, l'objectif du FRBHM, la sélection de k points de référence va sélectionner pendant plusieurs itérations en contenant plusieurs fois le premier point de référence sélectionné. Celui-ci restera donc le point sélectionné jusqu'à ce qu'il disparaisse de la sélection. Si $k = \Omega(P)$, c'est-à-dire que tous les points de référence sont sélectionnés à chaque itération, le point choisi ne changera jamais. L'algorithme qui recalcule chaque point de l'historique du FVLI avec le FBCM est le FRBHM continu.

FRBHM continu

Le FRBHM continu, ainsi nommé car il effectue le passage du discret au continu dès la constitution des ensembles de points fournis à l'algorithme FVLI, fonctionne comme suit :

- comme le FRBHM basique, il sélectionne les K points de référence les plus proches de la mesure fournie ;
- pour chaque point sélectionné, les indices de Friis correspondant sont calculés et le FBCM est appliqué, utilisant les mesures fournies pour calculer les distances par rapport aux points d'accès ;
- l'algorithme FVLI est appliqué aux ensembles de points. Le résultat est la position estimée du terminal mobile.

L'algorithme 2.12 explicite les étapes du FRBHM continu. La différence avec le FRBHM prenant en compte la topologie est l'exécution du FBCM sur chaque point de l'ensemble *k_pts_list*. L'objectif du FRBHM continu est de passer le plus vite possible dans le domaine continu, de façon à augmenter la variabilité des distances dans l'algorithme FVLI. En effet, si l'on s'éloigne d'un point p_i , celui-ci a une probabilité importante d'être présent dans les K points sélectionnés pendant plusieurs itérations. De fait, s'il est présent dans tous les ensembles, il sera considéré par l'algorithme FVLI comme la position du terminal mobile car la distance minimale entre les N ensembles sera nulle en passant par p_i à chaque itération. Au contraire, en calculant une position légèrement différente à chaque fois, grâce au FBCM, les distances varient plus souvent et sont différentes de 0. Par conséquent, la position calculée ne sera pas figée par l'algorithme FVLI.

```

FRBHM(measure, fvli_list)
1  int nb_ap ← measure.size()
2  Array of nb_ap float list_idx
3  Array of kPoints k_pts_list ← ss_map(measure, 1)
4  k_pts_list ← ss_map(measure, k)
5  for int i ← 1 to K
6  do for int j ← 1 to measure.size()
7      do list_idx[j] ← friis_idx(k_pts_list[i], measure[j])
8      k_pts_list[i] ← FBCM(list_idx, measure)
9  queue(fvli_list, k_pts_list)
10 return fvli(fvli_list)

```

FIG. 2.12 – Algorithme FRBHM continu.

L'algorithme FVLI appliqué à des ensembles de points quelconques requiert cependant de calculer les distances entre les points. Nous proposons une représentation qui permet ce calcul en tenant compte de la topologie.

2.4.4 Calcul de la matrice des distances du FVLI dans le FRBHM continu

La prise en compte de la topologie nécessite sa modélisation pour l'utiliser dans l'algorithme FVLI pour le FRBHM continu. Nous proposons de décrire l'espace du lieu de mise en œuvre suivant deux types d'éléments :

- un ensemble E de sous-espaces convexes ;
- un ensemble C des points de connexion entre sous-espaces.

E décrit des sous-espaces convexes correspondant à des éléments de la topologie tels que les salles dans un bâtiment. Il est nécessaire que les sous-espaces soient convexes de sorte que tout trajet effectué à l'intérieur d'un sous-espace puisse être direct, c'est-à-dire décrivant une droite interne au sous-espace. L'ensemble C indique les points par lesquels des éléments de E communiquent. Les points de l'ensemble C représentent par exemple les portes d'un bâtiment. Le calcul de la matrice des distances entre les points de connexion se fait suivant l'algorithme décrit à la section 2.3.1.

Le calcul de la distance entre deux points se fait suivant l'algorithme de la figure 2.13. L'algorithme prend en paramètre les points p_1 et p_2 . Ce sont les points dont on cherche à déterminer la distance qui les sépare. La fonction *get_zone* permet de connaître le sous-espace d'appartenance d'un point. La fonction *distance* calcule la distance euclidienne entre deux points. La fonction *voisins* est la fonction qui vérifie que deux sous-espaces sont voisins, c'est-à-dire qu'ils ont au moins un point de connexion commun. La fonction *connexion* renvoie la liste des points de connexion communs aux deux sous-espaces qui lui sont passés en paramètre. La matrice D est la matrice des distances topologiques entre les points de l'ensemble C .

Le principe de l'algorithme est de chercher les zones respectives de p_1 et p_2 , respectivement z_1 et z_2 . Si $z_1 = z_2$, p_1 et p_2 appartiennent à la même zone donc leur distance est la distance euclidienne entre eux. Sinon, on cherche l'ensemble des points de connexité de z_1 et de z_2 . Les

```

DISTANCE(Point  $p_1$ , Point  $p_2$ )
1  int  $z_1 \leftarrow \text{get\_zone}(p_1)$ 
2  int  $z_2 \leftarrow \text{get\_zone}(p_2)$ 
3  if  $z_1 = z_2$ 
4    then float  $d \leftarrow \text{distance}(p_1, p_2)$ 
5  else if  $\text{voisins}(z_1, z_2)$ 
6    then  $d \leftarrow \text{distance}(p_1, \text{connexion}(z_1, z_2)) + \text{distance}(p_2, \text{connexion}(z_1, z_2))$ 
7    else  $d \leftarrow \infty$ 
8    Array of int  $\text{connexion\_}z_1 \leftarrow \text{connexion}(z_1, *)$ 
9    Array of int  $\text{connexion\_}z_2 \leftarrow \text{connexion}(z_2, *)$ 
10   for  $i \leftarrow 0$  to  $\text{connexion\_}z_1.\text{size}() - 1$ 
11   do for  $j \leftarrow 0$  to  $\text{connexion\_}z_2.\text{size}() - 1$ 
12     do float  $d\_temp \leftarrow \text{distance}(p_1, \text{connexion\_}z_1[i]) + D[i][j]$ 
13     float  $d\_temp \leftarrow d\_temp + \text{distance}(p_2, \text{connexion\_}z_2[j])$ 
14     if  $d\_temp < d$ 
15       then  $d \leftarrow d\_temp$ 
16  return  $d$ 

```

FIG. 2.13 – Algorithme du calcul des distances entre points quelconques.

distances entre points de connexité sont calculées *a priori*, de la même manière que les distances entre points de référence (cf. l'algorithme de la fig. 2.9). Soient respectivement C_1 et C_2 les ensembles de points de connexité respectivement de z_1 et z_2 . Les éléments de C_1 et C_2 sont respectivement C_{1i} et C_{2j} , avec $1 \leq i \leq \Omega(C_1)$ et $1 \leq j \leq \Omega(C_2)$. On recherche le chemin de p_1 à p_2 de distance minimale, c'est-à-dire tel que :

$$d = \min(p_1 c_{1i} + c_{1i} c_{2j} + p_2 c_{2j}), \forall (i, j) \in [1; \Omega(C_1)] \times [1; \Omega(C_2)] \quad (2.5)$$

L'algorithme de calcul des distances entre les points permet de construire la matrice des distances utilisée par l'algorithme FVLI. Quand la matrice des distances est calculée, le reste de l'algorithme FVLI s'applique aux N ensembles de K points construits de la même façon que pour le FRBHM basique. On constate que cet algorithme peut également remplacer l'algorithme de calcul des distances entre les points de référence. En effet, des points de référence inclus dans une zone auront une distance égale à la distance euclidienne entre eux. Lorsqu'ils sont dans des zones différentes, l'utilisation d'un ou de plusieurs points de passage nous ramène encore à l'algorithme 2.13. La nuance tient à la possibilité de calculer la matrice des distances entre points de référence *a priori* et de l'utiliser à chaque appel au FRBHM basique.

2.5 Expérimentation et comparaison

Pour connaître l'impact de nos contributions sur la précision de la localisation, nous avons mené des expérimentations. Celles-ci visent à évaluer l'erreur commise par les systèmes de géolocalisation testés. Nous testons plusieurs modèles de l'état de l'art ainsi que nos contributions. Cela nous permet de comparer les résultats toutes choses égales par ailleurs.

Cette section est organisée de la façon qui suit. Premièrement, nous définissons le protocole des tests. Deuxièmement, nous exposons un par un les scénarios de test et leurs résultats. Nous tirons les conclusions des observations que nous avons faites sur les résultats des tests. En particulier, nous identifions les limites d'application de chaque technique testée.

2.5.1 Contextes des expérimentations

Dans cette sous-section, nous définissons les contextes des expérimentations successives que nous avons effectuées. Dans un premier temps, nous définissons les conditions des expérimentations et le matériel employé. Nous y spécifions en particulier les points d'accès utilisés. Dans un second temps, nous décrivons la disposition des points d'accès et des points de mesure dans le bâtiment Numérica.

Conditions des expérimentations

Plusieurs expérimentations ont été effectuées au cours de la thèse. Elles ne se sont malheureusement pas déroulées dans des conditions strictement identiques. En effet, les matériels, en particulier les points d'accès, sont utilisés également à des fins pédagogiques et indisponibles une partie de l'année. De plus, il est fréquent que des appareils soient redéployés sur d'autres projets et donc déplacés.

Nous distinguerons 3 expérimentations distinctes. La première ne disposait pas de matériel capable de mesurer la puissance des signaux en une seule opération. Les mesures ont donc été effectuées par associations successives à chacun des points d'accès disponible. L'association permet de connaître la mesure de la puissance du signal reçu. Un script shell est suffisant pour obtenir les données. Bien qu'elle permette de mesurer les puissances des signaux transmis par les points d'accès, cette méthode n'est pas viable pour un déploiement réel car l'association périodique à des points d'accès rend impossible toute communication via Wi-Fi. En comptant des délais d'association entre le terminal et un point d'accès, on peut obtenir une mesure de puissance pour un point d'accès chaque seconde. Les caractéristiques des points d'accès sont les suivantes : le point d'accès Netgear émet avec une puissance de 100 mW (20 dBm), les points d'accès Cisco émettant avec une puissance de 30 mW (15 dBm). Tous les points d'accès disposent d'antennes à 2 dBi. Le terminal mobile dispose d'une carte Wi-Fi PC-Card de modèle Cisco Aironet 350.

La seconde expérimentation repose sur l'utilisation de Radio Tap (RTAP). Ce mode de fonctionnement de l'interface Wi-Fi encapsule les paquets reçus dans un nouvel en-tête. L'en-tête RTAP contient des données relatives à Wi-Fi, en particulier des indicateurs de puissance du signal du paquet. La mesure incluse à l'en-tête RTAP est correcte. Cependant, elle requiert pour le récepteur d'être en mode moniteur, c'est-à-dire uniquement à l'écoute des canaux de communication. C'est en revanche un mode plus intéressant car il permet d'obtenir plus de mesures plus rapidement que le script shell de la première expérience. En effet, on mesure la puissance des balises (*beacons*) émises par les points d'accès. Sachant qu'un point d'accès émet 10 balises par seconde et que l'on capte toutes les balises en permanence, on peut obtenir plusieurs dizaines, voire certaines, de mesures par seconde. Dans cette expérimentation, les points d'accès disponibles et connus sont un routeur sans-fil Linksys WRT54G, un point d'accès Netgear et un

point d'accès Cisco, dans un bâtiment voisin. Le point d'accès Cisco émet à une puissance de 30 mW tandis que les deux autres émettent à une puissance de 100 mW. D'autres points d'accès sont présents, mais leurs caractéristiques ne sont pas connues. Ils sont exploités par l'algorithme basé sur les points de référence et ignorés par le FBCM.

La troisième expérimentation repose sur des mesures effectuées sur l'infrastructure. Pour cela, nous avons utilisé des mini-PC équipés de cartes Wi-Fi Intel BG2200. Ces mini-PC, au nombre de 5, sont déployés pour remplir le rôle de points d'accès capables de mesurer les puissances des signaux. Les signaux Wi-Fi sont ceux des paquets envoyés par le terminal mobile. Les points d'accès ne considèrent que les paquets d'un format prédéterminé pour être une requête de localisation ou de calibration. Les coordonnées des 5 points d'accès sont connues et utilisables par tous les algorithmes, que ces derniers se basent sur des données de référence ou le calcul de la position du terminal mobile. Les 5 mini-PC ont des antennes aux gains identiques de 5 dBi.

Une quatrième expérimentation a vu le jour. En effet, tester les algorithmes Viterbi-like requiert d'observer un mobile en déplacement. Or, la grille de mesures établies pour les seconde et troisième expérimentations a lieu sans continuité du déplacement. Pour pallier ce manque, nous avons effectué un déplacement dans le bâtiment, lors duquel les mesures s'effectuent au cours du mouvement, sans que le terminal ne s'arrête ni ne requiert d'intervention de l'utilisateur pour effectuer les mesures.

Disposition des équipements et des mesures

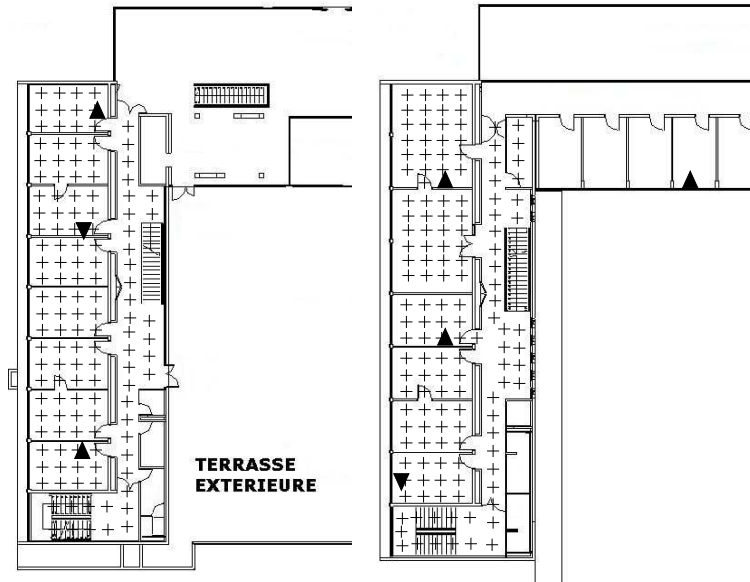


FIG. 2.14 – Disposition des points d'accès lors des expérimentations.

La figure 2.14 décrit la disposition des points d'accès lors de deux expérimentations. Les positions des points d'accès de la première expérimentation sont représentées par les triangles orientés pointes vers le bas. Les autres triangles, pointes vers le haut, sont les positions des points d'accès lors de la dernière expérimentation. Les croix sont les positions des points de mesure que

nous avons utilisés lors des seconde, troisième et quatrième expérimentations.

2.5.2 Protocole de test

Les expérimentations sont menées au rez-de-chaussée et au premier étage de Numérica. Nous disposons de 5 points d'accès et d'un ordinateur portable⁹.

Nous testons les projets suivants de l'état de l'art : SNAP-WPS [26], RADAR [15] et les travaux d'Interlink Networks [27]. Nos contributions testées sont le FBCM et les trois FRBHM. Dans certains cas, les systèmes FRBHM et RADAR sont testés sans intervention des algorithmes de type Viterbi. En effet, pour utiliser les algorithmes de type Viterbi, il faut que les mesures et les demandes de localisation soient effectuées dans un intervalle de temps et au cours du déplacement d'un individu.

La collecte de données pour les expérimentations s'effectue selon deux méthodes. La première méthode consiste à faire des mesures dans toutes les salles, à intervalle d'un mètre. Ceci permet d'obtenir une grille dense de points. L'ensemble des points mesurés pourra servir de cartographie des puissances et de points de calibration. Cet ensemble de points est l'ensemble S_1 . La densité des mesures permet de tester plusieurs densités inférieures ou égales. La seconde méthode consiste à se déplacer avec le terminal mobile en mesurant les puissances des signaux reçus. Cette méthode a pour objectif de pouvoir évaluer l'impact des algorithmes de type Viterbi. L'ensemble de points obtenus par cette méthode est S_2 .

Les mesures sont doubles : pour chaque point, une mesure est effectuée par l'infrastructure et une seconde mesure est effectuée par le terminal mobile. Cela nous permet d'étudier l'impact des acteurs de la mesure sur la précision de la localisation.

Pour évaluer la précision des systèmes de géolocalisation au cours des tests, nous enregistrons pour chaque point testé ses coordonnées réelles ainsi que les coordonnées définies par le système testé. Puis, nous calculons l'erreur de positionnement, déterminée par la distance euclidienne entre les coordonnées réelles du point et les coordonnées déterminées par le système de géolocalisation. Nous exposons dans la sous-section des résultats la moyenne des erreurs pour chaque système, dans les différentes conditions de test, ainsi que les écart types correspondants. Dans les courbes exposant les résultats, plus une courbe est basse, plus le système est donc précis.

2.5.3 Résultats et analyse

Dans cette section, les tests effectués sont détaillés. Nous y comparons les résultats de plusieurs systèmes de géolocalisation en fonction de la quantité de données disponibles. Les résultats obtenus sont discutés et permettent de mettre en lumière les points forts et les limites de chacun des systèmes. Non seulement nous proposons un test de nos contributions, mais de plus, nous étendons les expérimentations des travaux de l'état de l'art. En particulier, nous étudions l'impact de la quantité des données de référence et du nombre de points d'accès sur la précision de l'ensemble des systèmes testés. L'analyse permet d'envisager des scénarii de prédilection pour chaque système de géolocalisation.

⁹D'autres points d'accès pourront intervenir dans le processus de géolocalisation, mais ils ne seront pas utilisables pour tous les systèmes, car ils ne nous appartiennent pas et nous ne savons pas où ils sont situés.

Résultats

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM
1	8,99	4,47	16,33	5
2	8,99	5,57	16,2	4,51
3	8,99	6,15	11,57	5,3
4	8,99	7,78	5,92	5,87

TAB. 2.3 – Résultat des expérimentations, mesures effectuées par le terminal mobile.

Le résultat de la seconde expérimentation est décrit par le tableau 2.3. Il expose l'erreur commise par chaque modèle en fonction de la densité des points de référence. Les points de référence sont espacés de 1 mètre à 4 mètres. L'espacement de 4 mètres correspond à un point de référence par bureau, c'est-à-dire par zone homogène.

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM
1	13,81	3,16	11,88	4
2	13,81	3,44	11,38	4,28
3	13,81	6,61	13,83	4,33
4	13,81	3,84	9,8	4,7

TAB. 2.4 – Résultat des expérimentations, mesures effectuées par l'infrastructure en disposant de 3 AP.

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM
1	13,96	2,49	13,15	3,25
2	13,96	2,8	12,65	3,52
3	13,96	2,98	15,92	3,57
4	13,96	3,14	10,14	3,88

TAB. 2.5 – Résultat des expérimentations, mesures effectuées par l'infrastructure en disposant de 4 AP.

Les tableaux 2.4, 2.5 et 2.6 exposent les résultats de la troisième expérimentation. Les résultats proviennent de l'utilisation des données complètes, soit les mesures de puissance du signal sur 5 points d'accès, ou alors des données partielles, avec 3 et 4 points d'accès. Nous n'avons pas utilisé moins de points d'accès de façon à conserver la possibilité du calcul de la trilatération. Les résultats pour 3 et 4 points d'accès sont les moyennes de, respectivement, 10 tests et 5 tests, c'est-à-dire les combinaisons de, respectivement, 3 et 4 points d'accès parmi 5.

Les résultats de la quatrième expérimentation sont exposés dans les tableaux 2.7, 2.8 et 2.9. Tout comme pour la troisième expérimentation, les modèles de géolocalisation sont testés avec 3, 4 et 5 points d'accès. De plus, les modèles incluant la prise en compte de l'historique par un algorithme de type Viterbi sont également testés. Les algorithmes de type Viterbi ont été testés avec les paramètres K et N prenant pour valeurs respectives K et N .

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM
1	13,92	2,27	14,58	3,12
2	13,92	2,62	14,09	3,34
3	13,92	2,71	18,27	3,34
4	13,92	3	10,73	3,73

TAB. 2.6 – Résultat des expérimentations, mesures effectuées par l’infrastructure en disposant de 5 AP.

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM	RADAR + Viterbi	FRBHM + topologie	FRBHM continu
1	12,41	5,54	9,78	5,94	5,43	6,1	6,37
2	12,41	5,69	9,13	6,19	5,49	6,07	6,54
3	12,41	5,88	11,09	6,23	5,73	6,2	6,63
4	12,41	5,83	7,01	6,47	5,87	6,45	6,74

TAB. 2.7 – Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l’infrastructure en disposant de 3 AP.

Analyse

Plusieurs conclusions apparaissent à l’étude des résultats. Premièrement, la seconde expérimentation est la seule qui permette au FRBHM basique d’être plus précis que RADAR. Dans les expérimentations basées sur des mesures effectuées par l’infrastructure, RADAR est le système le plus précis. Bien que nous pensions accroître la précision du modèle en lui ajoutant une étape de trilatération, il semble que cela ne fonctionne pas toujours. Cependant, nous n’avons testé que des densités de points de référence allant de 1 mètre à 4 mètres entre deux points de référence. Il est possible qu’une densité inférieure dégrade plus rapidement les résultats du système RADAR et que le FRBHM devienne le plus précis.

Deuxièmement, nous constatons sans surprise que la précision du système proposé par Interlink Networks [27] ne varie pas en fonction de la densité des points de référence. L’explication de ce phénomène est évidente : IN n’est pas calibré par des mesures *a priori*, la variante de la formule de Friis utilisée étant toujours la même.

Troisièmement, nous remarquons un comportement étonnant du FBCM. Sa précision ne se

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM	RADAR + Viterbi	FRBHM + topologie	FRBHM continu
1	11,98	5	10,21	5,47	4,93	5,51	5,74
2	11,98	4,9	9,6	5,44	4,85	5,2	5,65
3	11,98	5,45	12,25	5,62	5,23	5,33	5,92
4	11,98	5,29	6,93	6,03	5,18	5,8	6,32

TAB. 2.8 – Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l’infrastructure en disposant de 4 AP.

Maillage (m)	Interlink Networks	RADAR	FBCM	FRBHM	RADAR + Viterbi	FRBHM + topologie	FRBHM continu
1	11, 63	4, 74	10, 75	4, 95	4, 83	5, 08	5, 1
2	11, 63	4, 48	10, 1	4, 79	4, 55	5, 02	5, 06
3	11, 63	5, 26	13, 56	5, 09	5, 22	5, 2	5, 27
4	11, 63	5, 03	7	5, 94	4, 81	5, 8	6, 09

TAB. 2.9 – Résultat des expérimentations sur un terminal en mouvement, mesures effectuées par l’infrastructure en disposant de 5 AP.

dégrade pas systématiquement. Dans les expérimentations avec mesures par l’infrastructure, la précision tend à augmenter lorsque la densité de points de calibration passe de 1 tous les mètres à un tous les deux mètres. Puis, la précision chute lors du passage à un point de calibration tous les 3 mètres pour augmenter à nouveau quand il y a un point de calibration tous les 4 mètres. L’amélioration du FBCM dans la seconde expérimentation est même totalement inverse à la densité de points de calibration, s’offrant même le luxe de surpasser la précision de RADAR. De par la position des points de calibration dans chaque densité, nous expliquons cette situation par la proximité des points de calibration avec les bords des zones homogènes, où le signal est moins stable qu’au centre des zones homogènes.

Les modèles utilisant l’historique des positions et l’algorithme FVLI sont moins précis que leurs pendants sans historique. Bien que le FVLI ait été ajouté pour accroître la précision, il provoque l’effet inverse. En effet, lorsqu’un point est sélectionné par le FVLI, il fait partie des k plus proches de la mesure dans l’espace du signal. Cependant, ce n’est pas toujours le plus proche. Par conséquent, l’évaluation des distances entre ce point et les points d’accès varie plus du fait de la distance dans l’espace du signal. C’est pourquoi on remarque une baisse de précision par l’usage du FVLI.

Il apparaît globalement que les méthodes basées sur la cartographie des puissances sont plus précises que des méthodes axées sur la modélisation de la propagation du signal.

On remarque également l’influence du nombre de points d’accès sur la précision. À l’exception du FBCM et d’IN, la précision des systèmes de géolocalisation décroît en même temps que le nombre de points d’accès. Les comportements du FBCM et d’IN s’expliquent par le fait que, sur 5 points d’accès, certains sont mal placés et vont réduire la précision lorsqu’ils sont utilisés en conjonction avec les autres.

Conclusions et perspectives

Parmi les conclusions que nous avons tirées de nos expérimentations, nous remarquons premièrement que le FBCM et la calibration permettent d’adapter le modèle de propagation des ondes à un bâtiment en particulier. Cependant, le FBCM reste imprécis pour un usage dans une application contextuelle. On remarque par extension que les autres modèles basés sur la trilatération seule souffrent des mêmes lacunes en termes de précision. Cette observation s’explique par l’hétérogénéité de la topologie dans les bâtiments. En effet, les modèles basés sur la propagation des ondes et la trilatération supposent un modèle de propagation unique par point

d'accès. Or, la topologie rend cette supposition fausse. En effet, dans un bâtiment, la relation entre l'atténuation du signal et la distance qui sépare l'émetteur du récepteur n'est pas bijective. On ne peut donc pas en déduire une relation réciproque, à moins de faire une approximation. Comme nous l'avons vu dans le calcul des indices de Friis, cette approximation conduit à des erreurs importantes dans l'évaluation de la distance. L'erreur se répercute bien entendu sur la précision de la géolocalisation.

Deuxièmement, nous avons remarqué que les modèles basés sur une cartographie des puissances voient leur précision augmenter avec la densité des points de référence. Par conséquent, ils demandent du temps pour être déployés. Ils sont également moins dynamiques. L'usage de l'algorithme de type Viterbi permet d'accroître la précision en supprimant les ambiguïtés entre points proches dans l'espace du signal mais éloignés dans l'espace cartésien. L'algorithme FVLI permet une meilleure précision car il prend en compte la topologie en considérant les déplacements possibles.

L'atout de l'algorithme FVLI en espace continu par rapport l'algorithme de type Viterbi basique tient à la discrétisation de l'espace pour le second. Nous avons remarqué que l'algorithme de type Viterbi tend à sélectionner le même point de référence sur plusieurs itérations successives. Le passage de l'espace, discret, des points de référence à l'espace, continu, cartésien pallie ce problème. En effet, il n'existe plus de distances nulles entre des points appartenant aux ensembles constitués à chaque itération.

Le modèle hybride FRBHM basique a une précision de l'ordre d'une salle (3 mètres). Il suffit pour envisager des applications relatives au contexte. Le FRBHM continu et le FRBHM avec topologie sont moins précis, contrairement aux attentes. En effet, la principale conclusion du comparatif effectué est le manque de précision des modèles basés sur la propagation des ondes dans un environnement hétérogène. Dans les 3 formes de FRBHM, peu de données *a priori* sont requises :

- ▶ il faut une mesure par salle, voire une tous les 4 mètres dans le cas de grandes salles ou de couloirs ;
- ▶ une connaissance basique de la topologie : où sont les murs, les portes et les escaliers.

Actuellement, les FRBHM sont déployés manuellement, par un utilisateur formé qui peut renseigner toutes les données nécessaires. À terme, cependant, il sera possible d'envisager un nouveau modèle qui se déploiera et se calibrera tout seul. Nous avons déjà mené des ébauches d'expérimentations pour localiser des points d'accès en couplant un récepteur Wi-Fi et un GPS.

Nous pouvons également concevoir un système dans lequel les points d'accès mesurent leurs réceptions respectives et se localisent relativement les uns aux autres. Ils ont ensuite la possibilité de localiser les terminaux mobiles dans le repère qu'ils ont construit. Combiner ce système avec l'hybridation Wi-Fi/GNSS permet de déterminer les coordonnées géocentriques des points d'accès. Le changement de repère permet alors de connaître également les coordonnées géographiques des terminaux mobiles. Dans le cadre où l'on n'envisagerait pas l'hybridation avec un GNSS, on peut aussi recourir à une recalibration dynamique grâce à des balises RFID. Ces propositions sont en cours d'étude et devraient permettre d'améliorer la réactivité du système de géolocalisation.

Deuxième partie

Prédiction de la mobilité

Chapitre 3

Modèles d'apprentissage des transitions

Introduction

La mobilité que nous nous proposons d'étudier est la mobilité physique [32]. En effet, nous considérons que la mobilité physique est la cause de la mobilité au sein du réseau. En particulier, la sortie du terminal mobile de la zone de couverture d'un point d'accès est directement liée à sa mobilité physique et aura un impact sur la mobilité dans le réseau : la connexion avec le réseau sera interrompue.

La gestion de la mobilité est possible suivant deux approches opposées : l'approche réactive d'une part et l'approche pro-active d'autre part. L'approche réactive déclenche une résolution des problèmes quand ceux-ci ont été détectés. Malheureusement, dans le cadre du multimédia mobile, ce choix aura pour effet de faire perdre le service à l'utilisateur. Par opposition, l'approche pro-active cherche à détecter l'imminence du problème pour le résoudre ou l'éviter avant son apparition. Dans le cadre du multimédia mobile, cette approche permet de ne pas interrompre les services de l'utilisateur.

Nous avons par conséquent choisi de faire reposer la gestion de la mobilité sur l'approche pro-active. Pour prévoir les interruptions des services, il faut déterminer quand un terminal mobile changera de point d'accès au réseau filaire, voire quand il quittera ce dernier. Nous proposons de prédire la position des terminaux mobiles en nous basant sur leur observation et un modèle d'apprentissage. L'observation consiste à détecter et quantifier la mobilité de chacun des terminaux mobiles. Nous nous intéressons en particulier à étudier *a posteriori* des déplacements de terminaux mobile afin d'en tirer un modèle auquel comparer des déplacements ultérieurs.

La prédiction de la mobilité requiert de qualifier le futur de la prédiction. Nous qualifions de *court terme* les quelques secondes qui suivent une observation. Le *moyen terme* commence à la limite du court terme et se termine après quelques minutes. Le *long terme* est au delà du moyen terme. Nous ne débattons pas des valeurs prises par les "quelques secondes" et "quelques minutes" car les valeurs de ces bornes varient selon l'application considérée et le mode de déplacement du terminal mobile.

En effet, les solutions retenues ne seront pas les mêmes selon les propriétés du terminal. La première propriété est l'outil de géolocalisation utilisé par le terminal mobile. Sa précision,

ramenée à l'échelle des déplacements du terminal mobile, et sa fréquence de rafraichissement permettra de se baser soit sur un calcul de la trajectoire du terminal mobile, soit sur un système d'apprentissage en découpant l'espace en états entre lesquels se déplace le terminal mobile. Le mode de déplacement est en relation avec la précision de l'outil de géolocalisation. En effet, le mode de déplacement est ce qui va donner son échelle aux déplacements du terminal mobile. Par exemple, un piéton localisé par le FRBHM se positionne à 3 mètres près. Pouvant changer de direction brutalement, sa trajectoire est difficile à calculer, d'autant que les courbes de sa trajectoire sont de petites dimensions par rapport à une erreur maximale de 3 mètres entre deux rafraichissements du FRBHM. Par opposition, malgré une erreur atteignant les 5 mètres avec le GPS, un véhicule personnel pèse au minimum 1000kg, et circule à grande vitesse sur une trajectoire restreinte par le tracé de la route. L'erreur de 5 mètres est, relativement à la taille du véhicule et les changements de direction plus "doux", plus faible que dans le premier exemple et il est envisageable de calculer la trajectoire du véhicule.

L'équation de la trajectoire se dérive pour obtenir les coordonnées du vecteur vitesse en tout point de la trajectoire du terminal. La vitesse et la position actuelle permettent de calculer la position qu'aura le terminal mobile dans un délai très proche. Ce type d'application est utilisable surtout pour des mobiles¹⁰ dont la vitesse est élevée et l'inertie importante. Par exemple, les trajectoires d'un avion ou d'une voiture peuvent se calculer de la sorte. En revanche, un piéton peut changer de direction quasi-instantanément car sa vitesse faible le lui permet. Cette propriété, combinée aux erreurs importantes des systèmes de géolocalisation (1 à 5 mètres dans les meilleurs cas) rend inapproprié l'usage du calcul de trajectoire pour des terminaux mobiles portés par des piétons. De plus, nous nous intéressons plus à l'effet du déplacement du terminal mobile sur l'état de ses connexions. Par conséquent, nous nous intéressons plus à la présence d'un terminal mobile dans l'une ou l'autre des zones de couverture du réseau qu'à sa position exacte dans l'espace.

Dans ce chapitre, nous présentons premièrement les méthodes de calcul de la trajectoire d'un terminal mobile. Ce type de modèle s'applique à la prédiction de la mobilité dans le court terme. Deuxièmement, nous présentons les modèles de prédiction qui ciblent le moyen terme. Ceux-ci fonctionnent selon l'apprentissage par le modèle de transitions observées et stockées.

3.1 Définition des horizons temporels

Dans cette section, nous définissons les horizons temporels que nous distinguons. Selon l'horizon envisagé, les techniques de prédiction de la mobilité utilisées diffèrent. De plus, la perception des horizons temporels est inféodée à la modalité du déplacement du terminal mobile. Nous définissons dans un premier temps la notion de modalité de déplacement. Dans un second temps, nous décrivons le court terme. Dans un second temps, nous présentons le moyen terme. C'est l'horizon temporel auquel nous nous intéressons le plus dans le cadre de la continuité des services. Dans un troisième temps, nous présentons le long terme, l'horizon temporel le plus difficile à étudier. Pour chaque horizon temporel, nous quantifions la valeur des seuils temporels qui le définissent, dépendant de la modalité de déplacement.

¹⁰au sens de physique du solide

3.1.1 Modalité de déplacement

La **modalité de déplacement** d'un terminal mobile est le moyen employé par ce dernier pour se déplacer. Chaque mode de déplacement a des propriétés de vitesse, d'accélération et de manœuvrabilité propres. Nous citerons notamment les modes suivants :

- le déplacement à pied ;
- les véhicules routiers ;
- le train ;
- l'avion.

Mode	Vitesse	Accélération	Manœuvrabilité
Piéton	4 à 6 km.h ⁻¹	instantanée	totale
Véhicule routier	30 à 130 km.h ⁻¹	variable	selon vitesse
Train grande ligne	plus de 100 km.h ⁻¹	longue	suit une voie
Avion (de ligne)	quelques km.h ⁻¹ au sol +700 km.h ⁻¹ en vol	n/a longue	faible faible

TAB. 3.1 – Quantification des propriétés des modes de déplacement.

Les propriétés de ces modes sont données dans le tableau 3.1. Le point commun entre les déplacements en véhicule est la vitesse élevée. Celle-ci donne une inertie (énergie cinétique) importante aux véhicules et leurs passagers, les rendant moins manœuvrables que des piétons. Ces derniers sont différents car ils sont beaucoup plus légers et lents et changent par conséquent facilement de trajectoire. L'accélération des véhicules n'est pas instantanée et se différencie là aussi du piéton qui atteint sa "vitesse de croisière" presque instantanément.

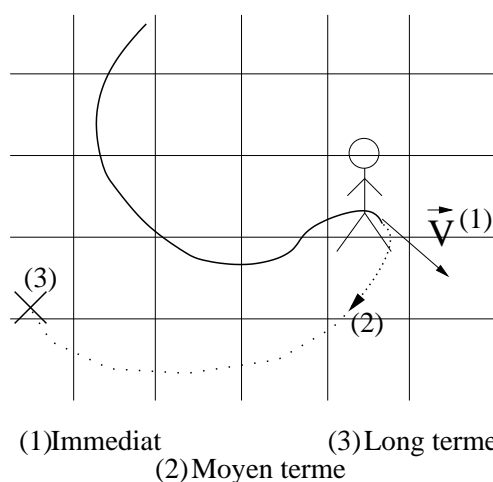


FIG. 3.1 – Horizons temporels de prédiction de la mobilité.

Les horizons temporels se définissent en fonction du moyen de pratiquer la prédiction. La figure 3.1 illustre le choix des horizons temporels. Le court terme est déterminé immédiatement

par la trajectoire du mobile. Le moyen terme est la prochaine position du mobile dans une portion de l'espace. Le long terme est la fin de la trajectoire du mobile.

3.1.2 Court terme

Dans nos travaux concernant la prédiction de la mobilité, le court terme est l'horizon temporel des déplacements internes à un état du système utilisé dans notre représentation. Dans le cas d'un réseau cellulaire (GSM) ou défini par des zones de couverture et des routeurs d'accès au réseau (802.11), il s'agit de l'étude de la trajectoire du terminal mobile au sein d'une cellule du réseau. Cette étude est rendue possible par la mise en équation du mouvement du terminal mobile de façon à calculer sa position future par variation de la variable du temps. La limite du court terme est dépendante de la vitesse de déplacement du terminal mobile ainsi que de sa modalité de déplacement. Dans le cas d'un véhicule, on considère que le court terme est une prédiction jusqu'à 3 secondes dans le futur. Dans le cas du piéton, le cas est plus complexe et dépend fortement du système de géolocalisation qui permet le suivi du terminal mobile. Si l'erreur du système de géolocalisation est inférieure à 50 cm, on peut envisager de prédire le mouvement du terminal mobile sur la prochaine demi-seconde. Si le système de géolocalisation est trop imprécis, aucune trajectoire du terminal mobile ne se calcule et toute prédiction de la mobilité à court terme est impossible.

3.1.3 Moyen terme

Le moyen terme est l'horizon temporel nécessaire à déterminer la prochaine localisation du terminal mobile dans le système défini par des cellules ou tout autre ensemble discret d'états. Dans tous les cas, on ne peut donner de valeur au temps nécessaire pour deux raisons. Premièrement, le temps passé dans un état dépend du dimensionnement de l'état. Deuxièmement, le terminal mobile peut stationner dans l'état qu'il occupe. Le moyen terme doit être traité par des méthodes statistiques en ce qui concerne le temps de présence dans l'état courant.

3.1.4 Long terme

Le long terme est le plus difficile à traiter. Il consiste à identifier des états attracteurs dans le système observé. En effet, il existe des états dans lesquels le terminal mobile sera amené à passer régulièrement. De façon moins formelle, il est évident que le sens porté au long terme est en relation avec les habitudes des usagers du réseau. Il est presque certain qu'un individu qui part de chez lui le matin y retournera en fin de journée. La problématique du long terme est de représenter les individus de façon à détecter les comportements liés à l'habitude. Le long terme est fortement lié à la sociologie mais d'autres systèmes peuvent néanmoins suivre des schémas prévisibles, comme des éléments mobiles ayant besoin de charger leur source d'énergie régulièrement.

3.2 Prédiction des trajectoires

La prédiction de la mobilité à court terme s'appuie sur l'étude des trajectoires empruntées par un terminal mobile. Dans cette section, nous présentons les manières pour obtenir les

informations de trajectoire et comment on les exploite pour établir une prédiction de la mobilité.

Les paramètres minimaux pour caractériser la mobilité sont la position, la vitesse et la direction du mouvement [33]. Divers techniques et appareils permettent d'obtenir ces informations. Dans [34], la prédiction de la mobilité est requise pour prédire la durée de connexion entre deux terminaux mobiles. L'objectif est d'optimiser le routage d'un réseau de terminaux Wi-Fi Ad hoc, comportant un GPS. Les données GPS des terminaux mobiles sont employées pour connaître la position, la vitesse et la direction du mouvement. Soient deux terminaux mobiles i et j . Leurs coordonnées respectives sont $(x_i; y_i)$ et $(x_j; y_j)$, leurs vitesses respectives sont v_i et v_j et leurs directions respectives sont θ_i et θ_j avec $0 \leq \theta_i, \theta_j \leq 2\pi$. De cette façon, la durée de connexion entre i et j peut être exprimée :

$$D_t = \frac{-(ab + cd) + \sqrt{(a^2 + c^2)r^2 - (ad - bc)^2}}{a^2 + c^2}$$

où

- ▶ $a = v_i \cos \theta_i - v_j \cos \theta_j$;
- ▶ $b = x_i - x_j$;
- ▶ $c = v_i \sin \theta_i - v_j \sin \theta_j$;
- ▶ $d = y_i - y_j$

Dans l'obtention de la trajectoire, un écueil important reste la fiabilité des données de position sur lesquelles on se base pour le calcul de position et l'extrapolation d'une équation. Les erreurs sont inévitables et peuvent être traitées par des processus tels que les filtres de Kalman [35] ou le *Double Exponential Smoothing* [36].

Dans [37], Bahl et al. présentent un modèle hiérarchique de mobilité. Le modèle local de mobilité utilise le calcul d'équations de trajectoire pour effectuer la prédiction à court terme. Le modèle global de mobilité considère les patrons de mobilité des usagers (*User Mobility Pattern*, UMP). La prédiction est effectuée par la corrélation entre le chemin actuel de l'utilisateur (*User's Actual Path*, UAP) et les UMP stockés. Le modèle global est motivé par le fait que les terminaux mobiles adoptent des cheminements, sinon identiques, du moins similaires. Le modèle local quant à lui est né de la constatation que les transitions inter-cellules semblent aléatoires mais ne le sont pas quand on observe de près les mouvements des terminaux mobiles. L'idée du modèle hiérarchique est de combiner les deux observations pour gagner en précision dans les prédictions immédiates grâce au modèle local, qui permet également d'accroître la précision du modèle global. La prédiction locale est effectuée en deux étapes :

- ▶ estimer l'état dynamique du terminal mobile en s'appuyant sur des mesures du signal ;
- ▶ sélectionner la cellule voisine la plus probable relativement à la forme de la cellule et à l'état dynamique du terminal mobile.

La prédiction locale proposée par Bahl et al. consiste à utiliser les informations de position et de direction les plus fines pour déterminer la trajectoire du terminal mobile et en déduire sa position à court terme. Deux étapes sont nécessaires pour effectuer la prédiction locale :

- ▶ estimer l'état dynamique du terminal mobile, donc sa trajectoire, en se basant sur ses mesures de RSSI ;

- sélectionner la cellule voisine qui intercepte la trajectoire du terminal mobile.

La première étape est effectuée à l'aide d'un filtre adaptatif. Les auteurs choisissent un filtre de Kalman modifié. Quand la trajectoire est calculée, son intersection avec l'un des bords de la cellule est résolue. Les cellules sont considérées comme étant des hexagones, chaque cellule ayant par conséquent au plus 6 voisins.

Des travaux concernant la prédiction de la mobilité appliquée à l'anticipation du handover dans un réseau IPv6 proposent de calculer la trajectoire du terminal mobile à partir de ses deux dernières positions [38]. Si $P_0(x_0; y_0)$ et $P_1(x_1; y_1)$ sont les deux dernières positions observées, les coordonnées de la trajectoire du terminal sont les suivantes :

$$x(t) = x_1 + t \times (x_1 - x_0) \quad y(t) = y_1 + t \times (y_1 - y_0) \quad (3.1)$$

Partant de cette équation, on considère pour chaque point d'accès une portée uniforme égale à G et les coordonnées de sa position $(a; b)$. Alors, l'équation de la limite de portée du point d'accès est :

$$(x - a)^2 + (y - b)^2 = G^2 \quad (3.2)$$

Avec les deux équations, il est possible d'écrire l'équation :

$$At^2 + Bt + C = 0 \quad (3.3)$$

où

- $A = \|\vec{U}\|^2$
- $B = 2 \times \overrightarrow{MO} \cdot \vec{U}$
- $C = \|\overrightarrow{MO}\|^2 - G^2$

La résolution de cette équation permet de déterminer trois scénarios possibles. Premièrement, si le déterminant de l'équation est inférieur à zéro, le terminal ne sera pas à portée du point d'accès. Si le déterminant est égal à zéro, sa trajectoire est tangente à la couverture du point d'accès. Si le déterminant est positif strictement, le terminal se dirige vers le point d'accès. En choisissant le point d'accès pour qui la valeur de t est la plus faible, on a de grandes chances que le terminal mobile reste longtemps dans la zone de couverture de celui-ci.

La prédiction à court terme s'appliquera dans le cas d'un véhicule positionné par GPS par exemple. Nous nous intéressons plus aux méthodes de prédiction à moyen terme dans le cadre de réseaux de terminaux mobiles Wi-Fi en intérieur.

3.3 Modèles d'apprentissage états-transitions

Les modèles d'apprentissage composés d'états et de transitions permettent de modéliser des systèmes discrets. Dans le cadre de systèmes continus, ces modèles requièrent de discrétiser le système décrit afin de procéder à l'observation et à l'apprentissage. Concernant la gestion de la mobilité, l'apprentissage nécessite d'établir un modèle formel suffisamment compact mais précis pour être utilisé par des équipements réseau.

3.3.1 Modèles de Markov

Les méthodes de prédiction discrète, adaptées au moyen terme, reposent sur l'apprentissage de données antérieures. Plusieurs éléments sont requis pour l'apprentissage. Premièrement, il est nécessaire de connaître les données antérieures, et donc d'identifier quelles propriétés sont pertinentes. Cette étape permet la journalisation des informations nécessaires à l'apprentissage. Deuxièmement, il est nécessaire de construire un modèle formel qui représente de manière compacte le résultat de l'apprentissage. Troisièmement, le modèle résultant de l'apprentissage est comparé à des observations en temps-réel pour prédire un comportement futur probable. Les modèles d'apprentissage utilisés dans la prédiction de la mobilité trouvent leur source dans la prédiction d'autres phénomènes, en particulier la prédiction de chargement de pages web. Ainsi, nous parlerons de prédiction de comportement plutôt que de prédiction de la mobilité, même si notre objectif direct est la prédiction de la mobilité.

L'approche la plus triviale se base sur un modèle de Markov (MM) dans lequel chaque état du modèle est calqué sur un état physique. Soit Φ l'ensemble des états physiques $\Phi_i, i \in \mathbb{N}$ du système dont on cherche à prédire les états futurs. Soit E l'ensemble des états $E_i, i \in \mathbb{N}$ du modèle de Markov représentant le système. L'ensemble E représente l'ensemble Φ de façon que E_i représente Φ_i . L'état physique peut être l'accès à un fichier, la présence d'un terminal mobile dans une zone, la visite d'une page web [39], c'est-à-dire tout événement ponctuel et discret. Le modèle de Markov est défini par le triplet (E, T, λ) où E est l'ensemble des états du modèle, T est la distribution de probabilité des transitions d'un état à un autre et λ est la distribution de probabilité initiale d'états de E . La distribution des probabilités de transition T est représentée par une matrice carrée de dimension N où N est le nombre d'états présents dans E , $N = \Omega(E)$. Chaque élément t_{ij} de la matrice de transition définit la probabilité de transition de l'état E_i à l'état E_j . La somme des probabilités sur une ligne ou une colonne de la matrice est égale à 1. On détermine t_{ij} par le calcul suivant : $t_{ij} = \frac{T_{ij}}{\sum_{k=1}^N T_{ki}}$ où T_{ij} est le nombre de transitions de l'état E_i vers l'état E_j . Le système étant dans un état Φ_i , on prédit l'état suivant comme étant l'état Φ_j correspondant à l'état E_j tel que la probabilité de transition de E_i vers E_j soit maximale.

	Séquence
1	P_1, P_2, P_3, P_4
2	P_1, P_2, P_3, P_5
3	P_1, P_2, P_4, P_6
4	P_1, P_2, P_4

TAB. 3.2 – Exemple de séquences d'états physiques pour l'apprentissage.

Dans le tableau 3.2, quatre séquences d'états physiques du système sont présentées. Ce sont les données d'apprentissage. La séquence n°1 par exemple consiste à rentrer dans le système par l'état P_1 , puis aller dans l'état P_2 , puis P_3 et enfin P_4 , avant de sortir du système. La figure 3.2 est le modèle de Markov de premier ordre construit à partir du tableau 3.2. Sur cette figure, les transitions entre les états sont étiquetées par la probabilité d'être empruntées. On y voit deux états qui n'apparaissent pas dans le tableau : S et F . Ces états sont respectivement l'état initial du modèle, celui dans lequel on ne connaît pas l'état physique actuel du système observé, et

l'état final, lorsque le système observé est à nouveau dans un état qu'on ne peut observer.

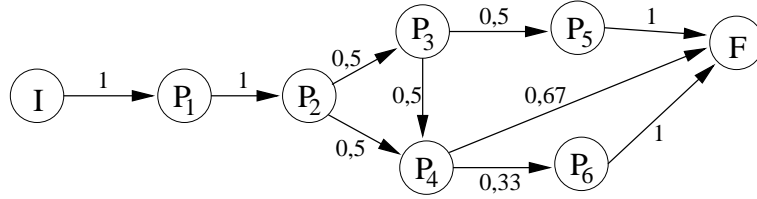


FIG. 3.2 – Modèle de Markov du premier ordre.

Ce modèle est simple à mettre en œuvre. Son inconvénient est de ne tenir compte que de l'état actuel pour déduire l'état suivant. Si l'on étudie les données du tableau 3.2 et le modèle de la figure 3.2 qui en découle, on voit que, si l'on observe l'état P_3 puis l'état P_4 , le modèle donne deux états successifs F et P_6 équiprobables. Cependant, d'après les données d'apprentissage, la séquence P_3 puis P_4 ne peut être suivie que par l'état F . D'autres travaux ont permis d'adapter le modèle de Markov à un historique plus long. Bien que le modèle de Markov ne puisse pas intrinsèquement déduire des transitions en fonction de $n \geq 2$ états antérieurs, il est possible de considérer un historique plus long en faisant porter à chaque état une information d'historique. En effet, bien qu'on considère un historique d'états physiques supérieur à 1, au sein du modèle, on ne se basera toujours que sur l'état courant pour déterminer son successeur.

Les degrés supérieurs à 1 du modèle de Markov sont construits à partir d'une succession de N -grams [40]. Un N -gram est une succession de N éléments d'états physiques du système observé. Un 2-gram, aussi appelé bigram, représente une transition. Un N -gram représente $N - 1$ transitions. Le 2-gram est utilisé pour établir le modèle de Markov du premier ordre. Un N -gram permet de construire le modèle de Markov d'ordre $N - 1$. Si l'on s'intéresse aux données du tableau 3.2 et à la construction du modèle de Markov du second ordre, un 3-gram possible (ou trigram) est P_1, P_2, P_3 . Dans le modèle de Markov du second ordre, on en déduit la transition de l'état $(P_1; P_2)$ vers l'état $(P_2; P_3)$.

Le modèle de Markov du k -ième ordre (k -th Markov Model, KMM) étiquette chaque état E_i du modèle par une succession d'états physiques $\Phi_{i_1}; \Phi_{i_2} \dots \Phi_{i_k}$. La succession d'états physiques représente le parcours de Φ_{i_1} , puis Φ_{i_2} , etc., jusqu'à Φ_{i_k} . Ce modèle est nommé modèle de Markov du k -ième ordre car, bien que l'absence de mémoire d'un processus markovien soit respectée, une profondeur d'historique de k états antérieurs est traitée. Le reste du modèle fonctionne comme le modèle de Markov du premier ordre. Dans la figure 3.3, les modèles de Markov d'ordre 2 et 3 issus des données du tableau 3.2 sont présentés. On y observe une probabilité de 1 pour la transition de $(P_3; P_4)$ vers F . Il y a donc un accroissement de la précision du modèle.

Afin d'augmenter la précision de la prédiction, il est possible de combiner des modèles de Markov d'ordres différents [41]. La méthode consiste à produire les modèles de Markov d'ordre 1 à K suivant les données d'apprentissage. Puis, dans la phase d'exploitation du modèle, la prédiction est effectuée par une recherche de la première prédiction possible en partant de l'ordre K et en allant vers l'ordre 1, en décrémentant l'ordre de 1 à chaque tentative de prédiction qui échoue. Cette méthode permet de tenir compte d'un historique de 1 à K états antérieurs. L'inconvénient est le cumul de plusieurs modèles de tailles importantes. Cette méthode est appelée *All-Kth*

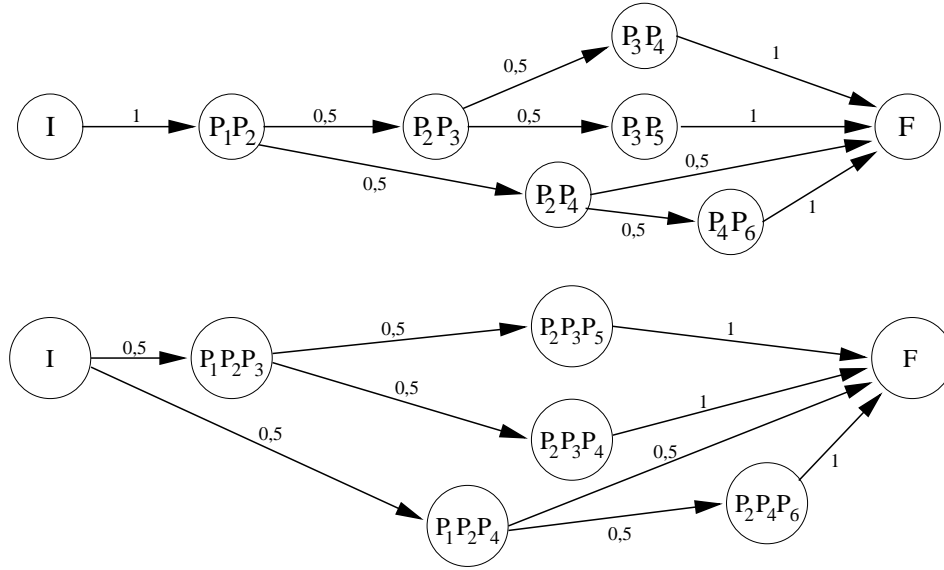


FIG. 3.3 – Modèles de Markov du second et troisième ordres.

Markov Model (AKMM).

L'augmentation de la taille d'historique prise en compte permet *a priori* d'augmenter la précision du modèle de prédiction. Cependant, nous verrons dans les expérimentations que ce n'est pas toujours le cas. L'inconvénient de l'augmentation de la taille d'historique est l'explosion du nombre d'états du modèle. Considérons un système pouvant se trouver dans N états physiques différents. Le modèle de Markov du premier ordre contiendra au maximum $N + 2$ états, du fait de l'ajout des états F et S . Le modèle de Markov du second ordre peut nécessiter $N \times (N - 1)$ états, le modèle de Markov de troisième ordre peut nécessiter $N \times (N - 1)^2$, et ainsi de suite. De nombreuses recherches ont donc été menées pour réduire le nombre d'états du modèle tout en conservant une précision acceptable.

Une méthode se base sur un traitement des données avant la construction du modèle [41]. L'objectif consiste à trouver les plus longues sous-séquences d'états se répétant (*Longest Repeating Subsequence*, LSR) dans les données d'apprentissage. Les autres données sont éliminées de l'apprentissage. De cette façon, le modèle de Markov construit nécessite moins d'états car il ne tient compte que des données les plus souvent présentes. Une sous-séquence est acceptable quand elle satisfait ces trois critères :

- une sous-séquence est composée de plusieurs états ;
- une sous-séquence apparaît plus que T fois dans les données, T étant typiquement égal à 1 ;
- bien qu'elle puisse être incluse dans d'autres LSR, la sous-séquence existe au moins une fois où elle est la plus longue.

Les auteurs de cette méthode, nommée *longest repeating subsequences*, considèrent que les transitions qui n'apparaissent qu'une seule fois dans les données d'apprentissage ne sont pas suffisamment pertinentes pour être conservées. Quand les LSR ont été extraites des données d'apprentissage, elles sont utilisées dans un modèle de Markov, créant ainsi le *All K^{th} -Order LRS*.

Plusieurs degrés de modèles de Markov sont testés.

	1-MM	1 st -order LRS	All-K th MM	All-K th -Order LRS
Model size (Ko)	372	136	8847	617
Nombre de transitions	13189	4953	217064	18549
Correspondances	25263	24363	25263	24363
Prédictions exactes	6402	6056	7704	6991

FIG. 3.4 – Results for LRS with various sizes.

Le tableau 3.4 présente les résultats des tests comparatifs entre les modèles de Markov d'ordres variés et le modèle basé sur les LRS. Les données d'apprentissage comportent à l'origine 25485 transitions. Le nombre de correspondances est le nombre d'états observés dans les données pour lesquelles un état a été trouvé dans le modèle. Une correspondance permet de faire une prédiction. La prédiction consiste à retourner l'état le plus probable parmi les successeurs de l'état courant, c'est-à-dire correspondant à l'observation courante. On remarque que les modèles basés sur les LRS sont un peu moins précis que les modèles de Markov qui emploient la totalité des données d'apprentissage. Cependant, la taille des modèles générés par les LRS est largement inférieure à la taille des modèles de Markov purs.

Dans [42], les modèles de Markov sélectifs (*Selective Markov Models*, SMM) sont présentés. Dans ceux-ci, le modèle *All-Kth Markov Model* est retraité *a posteriori* pour supprimer les états non pertinents du modèle afin de réduire sa taille. Les auteurs du SMM considère que des états ne sont pas pertinents dans 3 cas, correspondant à 3 types de suppression d'état :

- le modèle de Markov à suppression basée sur la fréquence (*Frequency-Pruned Markov Model*, FPMM) ;
- le modèle de Markov à suppression basée sur la confiance (*Confidence-Pruned Markov Model*, CPMM) ;
- le modèle de Markov à suppression basée sur l'erreur (*Error-Pruned Markov Model*, EPMM).

Le FPMM supprime les états des modèles d'ordre $k > 1$ qui n'apparaissent pas au minimum ϕ fois. ϕ est le seuil de fréquence (*Frequency Threshold*). Le choix de supprimer les états peu fréquents est dicté par le manque de fiabilité lorsqu'un état n'a été observé qu'un faible nombre de fois car il manquera de poids sur le plan statistique.

Le CPMM supprime les états dont les deux transitions sortantes les plus probables ont des probabilités trop proches pour prendre une décision. Le seuil de confiance (*Confidence Threshold*) se calcule grâce à la formule

$$\phi_c = p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

où n est la fréquence d'apparition de l'état et $z_{\alpha/2}$ est un paramètre qui, lorsqu'il augmente, conduit à supprimer plus d'états dans le modèle initial. Un état très fréquent reste pertinent même lorsque les probabilités de ses transitions sortantes sont proches. La présence de n dans le calcul du seuil de confiance permet de réduire ce dernier et de mieux conserver les états très fréquents.

L'EPMM prend en compte les erreurs commises par le modèle lorsqu'on l'applique à un échantillon de données indépendantes des données utilisées pour l'apprentissage. Cet échantillon

est l'ensemble de validation. L'EPMM fonctionne par comparaison d'un degré de modèle de Markov avec les degrés inférieurs. Un état est supprimé si le taux d'erreurs qu'il génère est supérieur au taux d'erreur le plus élevé de tous les modèles de degrés inférieurs. Deux niveaux sont envisagés pour l'EPMM. Le *Overall Error Pruning* (O-EPMM) teste toutes les sessions de l'ensemble de validation et enregistre pour chaque état les résultats des prédictions effectuées. Les résultats d'un degré de modèle de Markov sont comparés à ceux des modèles de degrés inférieurs, état par état. L'état (e_1, \dots, e_n) du modèle d'ordre n est comparé aux états (e_2, \dots, e_n) du modèle d'ordre $n - 1$, (e_3, \dots, e_n) du modèle d'ordre $n - 2$, etc. L'état est alors supprimé si son taux d'erreur est supérieur au plus haut taux d'erreurs des états des modèles de degrés inférieurs.

Dans tous les cas de suppression (FPMM, CPMM et EPMM), le modèle de Markov de premier ordre n'est pas modifié afin de conserver toutes les informations dans le modèle le plus simple. C'est également dans le modèle de premier ordre qu'on trouvera le plus d'états car les états du modèle de Markov de premier ordre sont des sous-états des modèles d'ordre supérieurs. Ceci signifie que, s'il est possible de trouver un état observé dans l'ensemble de tous les états de tous les modèles, on le trouvera au moins dans le modèle de Markov d'ordre 1.

3.3.2 Découverte des voisins

Le rapport de recherche n°5628 de l'INRIA[43] présente plusieurs algorithmes de découverte du voisinage des points d'accès dans un réseau 4G à couverture discontinue. Dans ce cadre, les points d'accès et les routeurs d'accès correspondent aussi bien à des équipements Wi-Fi qu'à des équipements d'autres normes (WiMax par exemple). Le contexte est un réseau dans lequel la partie sans-fil est connectée à la partie câblée au travers de contrôleurs d'accès (AC). Un ou plusieurs points d'accès (AP) sont connectés à chaque contrôleur d'accès.

IAPP (**I**nter **A**ccess **P**oint **P**rotocol) considère les relations de voisinage entre les points d'accès. Un terminal mobile qui quitte une zone de couverture d'un point d'accès va transmettre l'identificateur du point d'accès dont il a précédemment quitté la couverture. Ainsi, le second point d'accès ajoute le premier à son voisinage. De plus, il communique avec son voisin pour lui signifier la relation de voisinage. Connaissant ses voisins, un point d'accès peut préparer un handoff vers ses voisins. En effet, quand un terminal mobile quitte sa couverture, il communique le contexte du terminal à ses voisins. Ainsi, quand le terminal mobile pénétrera dans la zone de couverture d'un des points d'accès voisins, le service reprendra immédiatement.

CARD (**C**andidate **A**ccess **R**outer **D**iscovery) est un protocole fonctionnant sur le principe d'IAPP avec plus de paramètres. En effet, CARD part du principe que les points d'accès sont connectés à des routeurs d'accès¹¹, eux-même connectés au réseau câblé. Dans cette architecture, il est possible de connecter plusieurs points d'accès à un même routeur d'accès. Chaque routeur d'accès entretient une liste des correspondances entre les points d'accès et les routeurs d'accès de son voisinage. Un couple composé d'un point d'accès et d'un routeur d'accès fait partie du voisinage d'un routeur d'accès si le point d'accès a une couverture en intersection avec un de ses propres points d'accès.

CARD souffre d'un inconvénient car les communications se font lorsque le terminal mobile est

¹¹Ceux-ci sont similaires aux contrôleurs d'accès.

dans la zone de couverture conjointe de deux points d'accès voisins. Cette condition rend impossible l'utilisation de CARD dans les cas où les zones de couverture des points d'accès sont disjointes.

Par conséquent, les auteurs du rapport de recherche présentent NDP (**Neighbor Discovery Protocol**), leur propre algorithme de détermination du voisinage. Celui existe sous deux formes : distribué ou centralisé. Le fonctionnement de NDP est basé sur CARD. En effet, CARD est plus complexe que IAPP. Cependant, IAPP fonctionnant par rapport aux points d'accès, il est plus coûteux à mettre en œuvre que CARD. En effet, il y a plus de points d'accès que de contrôleurs d'accès. Dans sa version distribuée, les AC découvrent leur voisinage grâce à NDP en configurant chacun :

- une table d'AC candidats (CAC) au handoff. Cette table fait une correspondance entre les AP de l'AC courant et l'AC suivant ;
- une table (NAP) des AP et AC voisins (en correspondance) ainsi qu'un compteur de temps. Cette table permet à un AC de s'assurer qu'il est toujours un voisin des couples AC/AP de sa table NAP.

La version centralisée de NDP conserve l'ensemble des données des tables CAC et NAP au sein d'une seule table CAC contenant l'ensemble des données. Cette table est conservée par un serveur mandataire de mobilité.

À partir des informations des tables CAC et NAP, le déclenchement pro-actif d'un handoff se déroule quand un terminal mobile quitte la zone de couverture d'un point d'accès. Lorsque cet événement est déclenché, l'AC dont dépend l'AP recherche dans sa table CAC l'AC voisin et l'informe du contexte du terminal mobile.

3.3.3 Correction des erreurs d'observation

L'article [44] présente un modèle de prédiction des déplacements de terminaux dans les réseaux mobiles. Il s'appuie sur le modèle de Markov caché. La topologie du réseau est la suivante : des routeurs d'accès (Access Routers ou AR) sont disposés, chacun couvre une zone géographique que l'on abstrait en cellule. Si leurs zones de couvertures ont une intersection non vide, ils sont dits voisins. Les noeuds mobiles (Mobile Nodes ou MN) sont les terminaux mobiles qui parcourent le réseau composé par les AR. La prédiction de la mobilité des MN se base sur un apprentissage par les AR des mouvements des MN. Ce sont les MN qui se déplacent et communiquent l'AR de leur cellule d'origine à leur nouveau routeur d'accès quand il y a changement de cellule. Quand les informations lui sont transmises, l'AR les enregistre pour les exploiter. C'est un système d'apprentissage qui pondère les arcs de transition du modèle de prédiction de mobilité. Le modèle est basé sur le modèle de Markov à espace d'états caché (HMM).

Des simulations ont été effectuées par les auteurs de l'article. Les premières simulent des véhicules avec GPS indiquant la position des mobiles. Les secondes simulent l'utilisation de la distance par rapport à l'AR. Dans le second cas, la distance est obtenue par une méthode de calcul où la mesure de puissance est connue. Dans les deux types de simulations, les résultats concernant la précision du modèle sont positifs. Dans un autre article [45], des travaux additionnels ont été faits. Ceux-ci concernent notamment l'introduction d'observations bruitées, c'est-à-dire incluant une composante aléatoire. Les résultats obtenus sont moins précis. Ils res-

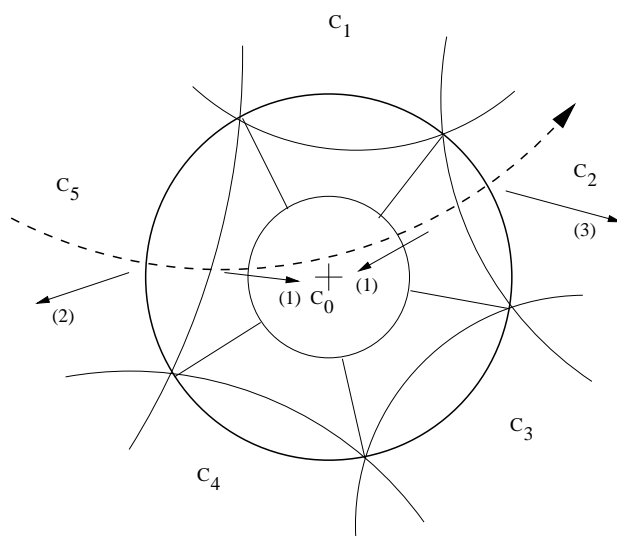


FIG. 3.5 – Utilisation des HMM dans la prédiction de la mobilité.

tent cependant acceptables, en particulier si l'on conserve les deux résultats les plus probables. Avec l'utilisation des deux premiers résultats, la prédiction reste fiable à 75% avec un bruit de 15% du signal et 5 observations utilisées pour établir la prédiction. Ces conditions sont le pire cas ayant été testé.

La contrainte majeure du modèle réside dans le fait que les tests sont réalisés uniquement dans des situations pour lesquelles la direction globale des mobiles est uniforme, en l'occurrence, toujours de la gauche vers la droite. Dans une problématique plus générale telle que la nôtre, cette hypothèse n'est pas vérifiée car les mobiles se déplacent dans toutes les directions. C'est ce qui explique que ce modèle ne peut pas être utilisé tel quel pour être appliqué à la problématique qui nous intéresse, et nécessite de ce fait une adaptation aux conditions de la prédiction de la mobilité.

3.3.4 Inférence Bayésienne

Les réseaux bayésiens ont été présentés dans la section A.5 où nous avons exposé qu'ils constituent un système de représentation des connaissances efficace et à même de calculer des probabilités conditionnelles. Il est possible d'intégrer beaucoup de variables dans un modèle bayésien. Dans le cas de la prédiction de la mobilité, des travaux exploitant les réseaux bayésiens ont été effectués [46].

La figure 3.6 représente le graphe des dépendances du réseau bayésien déployé. Il s'agit d'un réseau bayésien dynamique, c'est-à-dire qu'il est composé de tranches temporelles. Chaque tranche temporelle est définie par un modèle. Dans ce projet, les modèles de chaque tranche sont identiques. Des relations de dépendance existent entre les tranches temporelles.

Deux informations intéressent les auteurs du réseau bayésien : la position du terminal mobile et le temps de station à cette position. La position (CR pour *Current Room*) dépend d'une ou plusieurs positions précédentes, de l'heure dans la journée (TD pour *Time of Day*) et du jour de la semaine (WD pour *weekday*). Le temps de station à la position actuelle (CD pour *Current*

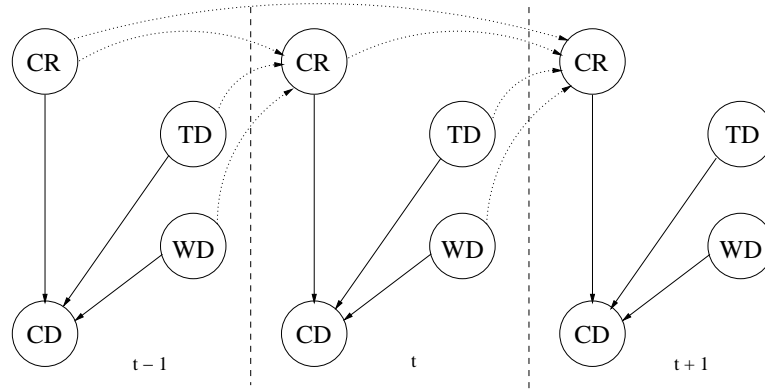


FIG. 3.6 – Graphe des dépendances du réseau bayésien de prédiction de la mobilité.

Duration) dépend de la position courante (CR), de l'heure de la journée (TD) et du jour de la semaine (WD).

Le modèle présenté exploite et traite plus de variables que les modèles présentés auparavant, qui s'intéressent uniquement à l'état suivant sans traiter le temps. Dans le modèle basé sur les réseaux bayésiens, on peut déterminer quelle sera la prochaine position du client mobile, mais également dans combien de temps il y parviendra. Cette donnée est extrêmement importante du fait de la problématique du handoff dans les réseaux de terminaux mobiles.

L'ensemble de fichiers utilisés pour les tests est constitué des données de 4 utilisateurs : A, B, C et D. Pour chaque utilisateur, deux périodes sont enregistrées : le printemps et l'automne. Il y a donc 2 ensembles de 4 fichiers, nommés *X_summer.data* et *X_fall.data* avec X prenant ses valeurs parmi A, B, C et D. Le contenu des fichiers est détaillé avec plus de précision dans la partie 4.5. Deux types de tests sont effectués sur ce réseau bayésien. Le premier type de tests est un test sans apprentissage préalable. Il consiste à commencer le test dès la première transition enregistrée dans les fichiers de trace. Le modèle se construit à mesure de la lecture. Le second type de tests consiste à construire le réseau bayésien avec les transitions d'un fichier de trace (*X_summer.data*) et de tester le réseau obtenu avec un fichier de trace différent (*X_fall.data*).

La quantification du résultat d'un test est le pourcentage de prédictions exactes. La prédiction est exacte quand la future position déterminée du terminal mobile est bonne. Dans la première série de tests, le délai avant changement de salle n'est pas calculé par le modèle bayésien. En revanche, quand aucune prédiction ne peut être effectuée, par manque de données préalables, l'étape du test est ignorée dans le résultat final. La précision du modèle est testée pour une séquence de salles comprise entre 1 et 5 salles précédentes lors de la tentative de prédiction. Les résultats complets sont disponibles dans le tableau 3.3. Alors que l'utilisation d'une salle précédente obtient une précision moyenne de 50 % de prédictions justes sur les 4 utilisateurs, augmenter la profondeur de l'historique diminue la précision (respectivement 45, 34, 27 et 14 % pour des profondeurs respectivement de 2, 3, 4 et 5 états précédents). Les auteurs expliquent ce fait par le manque de données d'apprentissage, donc la difficulté d'observer plusieurs fois la même séquence de déplacements.

L'imprécision du modèle est expliquée par le fait qu'une personne qui sort de son bureau peut se diriger n'importe où (transitions vers les autres salles proches de l'équiprobabilité). Les

Personne	Historique	1 salle	2 salles	3 salles	4 salles	5 salles
	A	52,81	48,89	34,85	28,18	11,85
	B	55,31	48,41	38,43	31,46	18,58
	C	42,64	39,24	26,32	22,76	9,96
	D	48,76	44,31	35,81	27,54	16,59

TAB. 3.3 – Modèle bayésien, résultat complet, sans apprentissage préalable.

auteurs proposent alors de ne tester que les cas où l'utilisateur n'est pas dans son propre bureau.

Les résultats présentés dans le tableau 3.3 ne tiennent pas compte du temps de présence de la personne dans la salle où elle est détectée, avant le passage vers une autre salle. Par conséquent, les auteurs de ces travaux proposent cette prise en compte via une modélisation statistique du temps passé dans chaque état pour l'ensemble des combinaisons des jours de la semaine avec le moment de la journée (choisi parmi matin, midi, après-midi et nuit).

	None	Time of day	Weekday	Both
Person A	51,17	48,37	38,36	29,31
Person B	58,49	55,7	50,05	42,65
Person C	72,15	68,2	65,1	53,95
Person D	52,07	47,47	42,64	33,41

TAB. 3.4 – Modèle bayésien, prise en compte du temps.

Les résultats obtenus sont donnés dans le tableau 3.4. On constate que la prise en compte du temps dans le modèle tend à dégrader la précision. Les auteurs de l'article imputent ce comportement à un trop faible nombre de données d'apprentissage. Cependant, dans le cas d'un système temps-réel, il est vital de prendre en compte le temps avant le déclenchement d'un événement.

Les auteurs traitent également un problème commun des systèmes d'apprentissage : la capacité à répercuter des changements d'usage. En effet, un réseau bayésien est long à répercuter un changement de comportement. Pour simuler ceci, 60 enregistrements de l'utilisateur A, suivis de 140 enregistrements de l'utilisateur C ont été fournis au réseau bayésien. Puis, deux tests ont été lancés : l'un avec 100 transitions apprises, le second avec 200. Dans le premier cas, 90 nouvelles transitions (90 %) sont nécessaires à la correction du modèle pour retrouver une précision semblable à celle préalable au changement. Dans le second cas, 130 nouvelles transitions (65 %) ont été nécessaires. La conclusion est qu'il n'est pas possible d'établir une règle donnant la quantité de nouvelles données à fournir au modèle pour qu'il redevienne précis, en fonction de la quantité de données qu'il a déjà intégrées.

3.3.5 Autres modèles

P. Bahl propose un modèle prédictif discret dans ses travaux sur la prédiction de la mobilité dans les réseaux ATM [37]. Ses travaux prennent place dans le contexte d'un réseau cellulaire dans lequel on souhaite prédire la future cellule où sera présent chaque terminal mobile de façon à

optimiser le service. Ce modèle est combiné au modèle présenté dans la section 3.2 pour effectuer une prédiction précise. Les pérégrinations des utilisateurs sont modélisées par les *User Mobility Patterns* (UMP), qui représentent une séquence de déplacements inter-cellulaires.

Un UMP de longueur n est de la forme $\{a_1, a_2, a_3, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n\}$ où les a_i sont des cellules. On considère de nouvelles pérégrinations, *User Actual Path* (UAP), comme étant des UMP “édités”. Les opérations d’édition des UMP autorisées sont les suivantes :

- insérer une cellule c en i^{ime} position d’un UMP donne l’UAP :
 $\{a_1, a_2, a_3, \dots, a_{i-1}, c, a_i, a_{i+1}, \dots, a_n\}$
- supprimer la cellule a_i en i^{ime} position d’un UMP donne l’UAP :
 $\{a_1, a_2, a_3, \dots, a_{i-1}, a_{i+1}, \dots, a_n\}$
- changer la cellule a_i en i^{ime} position d’un UMP par la cellule c donne l’UAP :
 $\{a_1, a_2, a_3, \dots, a_{i-1}, c, a_{i+1}, \dots, a_n\}$

Un coût est affecté à chaque opération, respectivement W_{Ic_i} , W_{Da_i} et W_{Cc_i} pour respectivement l’insertion, la suppression et la modification.

Les UMP sont classés par date d’apparition. Les dates sont des intervalles de temps pendant lesquels un UMP a été observé. L’objectif de la prédiction de la mobilité est de trouver pour un UAP observé durant un intervalle de temps donné, l’UMP qui le décrit le mieux. L’UAP étudié contient également, si elle existe, la prédiction locale déterminée par le calcul de trajectoire. La solution est obtenue par programmation récursive, en cherchant quel UMP coûte le moins pour devenir la même séquence que l’UAP. Le coût d’édition doit également être inférieur à un seuil t paramétré par l’utilisateur.

Conclusion

Dans cette section, nous étudions des modèles prédictifs. Tous ne sont pas issus de travaux relatifs à la mobilité. En effet, nous nous intéressons à la prédiction du moyen terme. Dans le moyen terme, nous partitionnons le territoire à couvrir en zones auxquelles des services seront associés. Par conséquent, nous nous plaçons dans un ensemble discret d’états. Les transitions entre les états s’étudient bien avec des modèles basés sur des états et des transitions, tels que les modèles de Markov. Les modèles probabilistes, réseaux bayésiens en particulier, sont également adaptés à la représentation d’événements discrets.

Cependant, les modèles étudiés souffrent de défauts à résoudre avant de pouvoir les mettre en application dans la prédiction de la mobilité. Les modèles de Markov étudiés ne prennent pas en compte la dimension temporelle. En effet, une page HTML peut être préchargée immédiatement. Ce n’est pas le cas des données multimédia riches, telles que les vidéos, qui occupent le réseau plus longtemps. Il faut par conséquent déclencher la politique de *handover* avec un délai pour réduire l’encombrement du réseau mais assez tôt pour ne pas rater la réalisation du *handover* par le terminal mobile. Le réseau bayésien dans le cadre de la prédiction de la mobilité tient compte de la dimension temporelle. Cependant, sa précision n’est suffisante que dans les cas où la prédiction est facile, c’est-à-dire quand on observe une transition probable à plus de 90%.

Aucun des modèles étudiés n’étant directement exploitable dans le notre cadre de travail, nous devons améliorer et adapter les modèles de Markov à l’anticipation du *handover*.

Chapitre 4

Prédiction de la mobilité dans les réseaux multimédia sans-fil

Introduction

Le contexte dans lequel nous nous plaçons est celui de la mobilité des terminaux dans un réseau Wi-Fi. Le terminal est localisé en se basant sur les mesures de puissance des signaux Wi-Fi reçus. Des services sont fournis au terminal en fonction de sa position. Les services sont fournis au terminal mobile lors de son déplacement. En conséquence, un terminal mobile peut, à tout moment, quitter la couverture de son point d'accès au réseau et en changer. Si c'est le cas et que rien n'est fait, l'utilisateur du terminal mobile voit les services interrompus. La gestion de la mobilité a pour objectif de pallier ce type de problème. Basculer d'un point d'accès du réseau à un autre est une procédure nommée *handover*. Le *handover* peut être soit "dur", c'est-à-dire qu'il y a une perte de connexion et une interruption, même minime, des services, soit "doux", c'est-à-dire que le basculement s'effectue sans interruption de service. Notre objectif est d'assurer un *handover* doux. Pour y parvenir, nous avons choisi de recourir à une méthode proactive. Nous basons notre méthode sur la prédiction de la mobilité afin d'obtenir un *handover* prédictif.

Sachant que les positions du terminal mobile au cours du temps sont connues, il est possible de le traquer et d'enregistrer ses déplacements. Exploiter ces données pour prédire les déplacements d'autres terminaux mobiles nécessite une modélisation. La problématique du modèle choisi est la suivante : *la prédiction de la mobilité consiste à représenter les données d'apprentissage de façon compacte sans perdre l'information pertinente ni surcharger les critères de modélisation et diluer les données dans des groupes de critères dont le poids statistique devient trop faible pour être fiable*. Répondre à cette problématique consiste à trouver l'équilibre entre complexité du modèle et masse de données conservées.

En effet, plus le modèle est complexe, plus les quantités de données regroupées seront petites. De même, plus le modèle est simple, plus les données perdent leur sens originel. De plus, plus la quantité de données maintenues dans le modèle est grande, plus le modèle est lourd à stocker et manipuler, posant un problème de vitesse d'exécution dans un contexte où la réactivité est critique.

Dans la suite de ce chapitre, nous présentons nos contributions en termes de prédiction de la mobilité. Premièrement, nous décrivons quelles données sont exploitées dans nos modèles et

nous donnons une structure récursive compacte permettant de les conserver. Deuxièmement, nous plaçons la problématique de prédiction de la mobilité dans le contexte de la diffusion de contenus multimédia, avec les spécificités temps-réel impliquées et les solutions potentielles. Troisièmement, nous proposons nos modèles d'apprentissage des pérégrinations, l'un basé sur les modèles de Markov, l'autre étant basé sur les réseaux bayésiens. Quatrièmement, nous décrivons la mise à l'épreuve des modèles pour en déduire leurs limites, en particulier en termes de trafic généré sur le réseau par le préchargement des contenus multimédia.

4.1 Problématique et choix

Nous nous plaçons dans le contexte de la prédiction de la mobilité à moyen terme. Le moyen terme de la prédiction se situe à un horizon de plusieurs secondes, le court terme étant le futur à moins d'une seconde et le long terme se jouant sur plusieurs heures/jours. Nous nous concentrons sur le moyen terme car il est certainement pertinent comparé au long terme, ce dernier ne permettant pas d'assurer la continuité des services en mobilité. Nous ne nous intéressons actuellement pas au court terme pour deux raisons :

- le court terme requiert la modélisation des trajectoires des terminaux mobiles, ce qui est difficile à obtenir d'un terminal mobile piéton, sujet à des changements radicaux de trajectoire. Ces changements sont trop difficiles à lisser, même avec des techniques comme les filtres de Kalman ;
- nous nous intéressons à des événements discrets qui mettent en jeu le changement de point d'accès au réseau d'infrastructure. Il s'agit d'un phénomène qui peut se représenter par l'étude de données d'apprentissage.

Le moyen terme nécessite d'observer les déplacements des terminaux mobiles dans leur usage courant. Les observations sont journalisées. La première problématique est de déterminer les données qui seront journalisées. Par extension, il est également nécessaire de savoir comment observer les déplacements des terminaux mobiles, question moins triviale qu'il n'y paraît. En effet, selon les objectifs envisagés lors du déploiement du système de gestion de la mobilité, ce ne seront pas les mêmes données qui seront pertinentes. Certaines données sont faciles à appréhender et à mesurer. C'est le cas par exemple de la connexion à un AP du réseau. D'autres données sont moins faciles à observer : par exemple, la position dépend de la précision du système de géolocalisation sous-jacent. La seconde source de difficulté dans la collecte des données pour l'apprentissage est la pertinence du découpage de l'espace des mesures en sous-espaces qui portent chacun une signification au sens des services.

Quand le découpage et/ou la nature des données journalisées ont été choisis, ce que l'on considérera comme acquis dans la suite de ce chapitre, une quantité importante de ces données est nécessaire pour que le modèle de représentation de la mobilité apprenne les schémas de déplacement des terminaux mobiles. La quantité de données doit être importante pour donner une masse statistique au modèle déployé. Dans le cas contraire, le modèle est en sous-apprentissage et ses résultats ne sont pas pertinents. D'une part, il est difficile de conserver toutes les informations journalisées dans le serveur chargé de la gestion de la mobilité. D'autre part, plus la quantité de données est importante, plus la recherche des données intéressantes pour un cas donné est longue. Il y a nécessité de représenter nos données de façon compacte. Compacter les données

induit de perdre certaines informations. L'objectif est de ne perdre que ce qui est le moins significatif pour garder les données qui permettent un recoupement précis avec les déplacements d'un terminal mobile en mouvement.

Notre contexte de la prédiction de la mobilité est celui d'un réseau composé de terminaux mobiles auxquelles sont délivrés des contenus multimédia. Les contenus sont délivrés aux terminaux par diffusion. La prise en compte de la mobilité est importante pour anticiper les interruptions dans les flux transmis aux terminaux. Une procédure de *handoff* doit être mise en place pour assurer la continuité des flux lorsqu'un terminal mobile est amené à changer de point de connexion à l'infrastructure du réseau. La seconde application que nous envisageons est l'anticipation des congestions dans le réseau. En particulier, dans certains réseaux de terminaux mobiles (par exemple, les réseaux 802.11), le canal de communication est partagé. Donc, plus le nombre de terminaux connectés à un point d'accès à l'infrastructure est important, plus le débit individuel des terminaux chute.

L'étude de la mobilité des terminaux mobiles intervient premièrement dans l'anticipation de la procédure de *handoff*. En effet, les nouvelles connexions doivent être préparées pour basculer de façon transparente vis-à-vis des applications et de l'utilisateur. De plus, certaines applications requièrent également des migrations de données, dans notre exemple, le cache des contenus multimédia doit migrer [47] lors du *handoff*. La prédiction de la mobilité intervient également dans le contrôle de la congestion du réseau. En effet, la prédiction de la mobilité de l'ensemble des terminaux mobiles du réseau permet d'en prévoir le nombre connecté à chaque point d'accès à l'infrastructure.

La gestion de la mobilité que nous proposons est fondée sur la prédiction des déplacements futurs des terminaux mobiles. Un simple traitement statistique ne suffit pas car les connexions sont "nominatives", c'est-à-dire que les connexions des terminaux ne sont pas interchangeables. Nous proposons de traiter les terminaux mobiles individuellement en utilisant l'étude des *sessions de déplacements*, aussi appelées *pérégrinations*, pour construire un modèle de type état-transitions. De la même façon, on pourrait reposer naïvement sur un traitement statistique du nombre de terminaux mobiles connectés aux points d'accès de l'infrastructure. Cependant, la prédiction de la mobilité traitant individuellement chaque terminal, elle permet de considérer plus finement les types de terminaux et les applications qu'ils utilisent. En effet, les contraintes varient d'une application à une autre. La diffusion de contenu multimédia doit être temps-réel et ne requiert pas une fiabilité à 100 % alors qu'un transfert de fichier peut souffrir d'un délai mais ne peut autoriser de perte de paquet.

Nous avons choisi un modèle discret pour deux raisons. Premièrement, considérer des points d'accès au réseau d'infrastructure revient à considérer des zones de couvertures connectés entre elles par leur proximité. Nous considérons comme état du modèle la présence dans une zone de couverture, c'est-à-dire la connexion au point d'accès qui couvre celle-ci. Deuxièmement, nous souhaitons un modèle qui s'applique à des terminaux mobiles portés par des piétons. D'une part un piéton peut facilement changer de direction lors de ses déplacements. D'autre part, la précision des méthodes de géolocalisation, sur lesquelles il est envisageable de travailler, n'est pas suffisante pour déduire une équation de la trajectoire du terminal mobile piéton. En effet, à la vitesse moyenne de déplacement d'un piéton, quelques mètres d'erreur de localisation rendent le calcul de trajectoire quasiment aléatoire.

Notre modèle doit également tenir compte du temps car l'anticipation du handoff doit être effectuée au moment pertinent. Si le handoff est prévu trop tard, la décision est inutile et les connexions du terminal mobile sont interrompues. Inversement, si le handoff est prévu trop tôt, des ressources réseau sont occupées inutilement et plus longtemps. De plus, il est possible qu'entre temps, les paramètres changent. Le terminal mobile peut être amené à se déplacer ailleurs, là où il aurait été possible de le prévoir si l'on avait attendu assez longtemps. Il peut également requérir de nouveaux services et ainsi invalider les données préparées pour le handoff.

Le contexte dans lequel nous nous plaçons est celui d'un terminal mobile dont nous avons observé les déplacements. Les observations sont journalisées et comparées à des données antérieures afin de déterminer statistiquement les déplacements potentiels du terminal mobile.

4.2 Journalisation et conservation des données

Étudier la mobilité des terminaux consiste dans un premier temps à observer et quantifier leurs mouvements. Dans cette section, nous énumérons premièrement les données que nous exploitons dans la prédiction de la mobilité ainsi que la raison de leur utilisation. Nous exposons également la représentation que nous donnons à ces informations pour les conserver.

4.2.1 Choix des données

Les données utilisées en apprentissage du système de prédiction de la mobilité doivent être définies précisément. Nous nous intéressons à un système de prédiction de la mobilité dans un espace d'états discrets.

Dans l'espace des déplacements des terminaux mobiles, nous définissons un état comme étant un sous-ensemble de l'espace où les terminaux mobiles se déplacent. Le choix du découpage est dépendant avant tout des possibilités de localisation des terminaux mobiles ainsi que de l'application du service de prédiction de la mobilité. Les deux découpages suivants en sont un exemple :

- la présence du terminal mobile dans la zone de couverture d'un point d'accès au réseau d'infrastructure ;
- le découpage selon les salles d'un bâtiment.

Le premier exemple est utile pour gérer le handoff du terminal mobile en cas de passage de la zone de couverture de son point d'accès actuel à la zone de couverture d'un point d'accès voisin. Le second exemple permet de préparer la diffusion de contenu multimédia adapté à la salle vers laquelle le terminal mobile se dirige. Il permet également, via une correspondance entre les salles et les zones de couverture des points d'accès déployés, de préparer un handoff du terminal mobile.

En nous basant sur la définition de l'état pour le modèle d'apprentissage, nous définissons maintenant la pérégrination. Une pérégrination est une succession d'états par lesquels le terminal mobile va transiter. Pour transcrire une pérégrination, nous enregistrons la succession des états auxquels le terminal mobile a été observé. De plus, la prédiction de la mobilité dans le cadre de la continuité de service répond à des contraintes de temps réel. Nous enregistrons donc également les dates de transitions.

Les informations définies ci-dessus sont fondamentales pour le fonctionnement de l'apprentissage. Elles peuvent être complétées par des données plus précises issues de capteurs. Par exemple, les mesures d'un accéléromètre ou d'un compas sont utilisables dans le cadre de la prédiction de la mobilité, en particulier en tant que variables ajustant les calculs de prédiction à court terme.

4.2.2 Problèmes potentiels et leurs solutions

Un modèle basé sur l'apprentissage des pérégrinations des terminaux mobiles peut être soumis à différents problèmes. Le premier problème est le suivant : que va-t'il se passer si une observation ne correspond à aucune pérégrination de l'échantillon d'apprentissage ? On peut choisir de ne rien faire, en particulier si aucun autre système de prédiction de la mobilité n'est applicable. On peut également effectuer une prédiction de la mobilité grâce à un autre outil, par exemple le calcul de trajectoire. Dans les deux cas, il peut être pertinent d'intégrer la donnée à l'échantillon d'apprentissage et de mettre à jour le modèle.

Plus généralement, il est possible que les usagers du réseau où est déployé le système de prédiction de la mobilité changent d'habitudes. On obtient alors une généralisation du premier problème. Il est alors nécessaire de ne rien faire pendant une période de réapprentissage ou de reposer sur un autre système de prédiction de la mobilité. Pendant ce temps, les nouvelles données sont enregistrées et substituées à l'échantillon d'apprentissage. Ce problème et sa résolution supposent d'être capables d'identifier un changement d'habitude sur l'ensemble de la population des usagers.

D'autres problèmes plus ponctuels peuvent apparaître, comme la défaillance d'un terminal mobile, mais nous ne les traitons pas car nous nous intéressons au fonctionnement général du système de prédiction de la mobilité.

4.3 Seuils de prédiction

Envisager la prédiction de la mobilité et la politique de handoff qui en découle peut être fait de deux façons. La première est naïve et consiste à réagir de manière binaire, c'est-à-dire en "tout ou rien". Le résultat de la prédiction pour chaque état potentiellement choisi par l'algorithme de prédiction est soit un handoff complet, soit aucun handoff. La seconde façon d'envisager la politique de handoff est de prendre une décision graduelle en fonction de la probabilité d'un état d'être le prochain. Dans le cas d'un service dont la qualité est quantifiable, il est possible de réduire la qualité pour améliorer l'efficacité du handoff.

Par exemple, dans le cas de la diffusion de contenus multimédia à encodage hiérarchique, on peut choisir de transmettre à un niveau de qualité déterminé par la probabilité de chaque état à être la cible de la prochaine transition [48].

4.3.1 Contexte de la continuité des services

Dans le contexte de la continuité de services, nous considérons en particulier la diffusion de contenus multimédia. Un contenu multimedia peut être vu comme un flux d'images que l'on transmet ou que l'on interrompt selon les circonstances. On peut également considérer que le flux est décomposable en plusieurs qualités, le débit requis par chacune étant proportionnel à

cette dernière [49]. On parle de flux hiérarchiques car, à un flux de base, s'ajoutent plusieurs flux supplémentaires qui accroissent la qualité de la vidéo.

Les services de diffusion multimédia, tels que la vidéo à la demande, la visioconférence et la téléphonie sur IP, ont également pour propriété d'être des transmissions en temps-réel. En effet, perdre une image d'une vidéo est irrémédiable, car toute image qui n'arrive pas en temps voulu dans la séquence est perdue. Il est par conséquent inutile de retransmettre une image non reçue par le client mobile. De plus, une image qui arrive trop tard perturbe la lecture de la vidéo. Il faut par conséquent assurer la transmission du flux en temps-réel. Dans ce cadre, la prédiction de la mobilité aide à anticiper les déconnexions du terminal mobile et à préparer le handoff vers le point d'accès suivant.

La prédiction de la mobilité dans un réseau de diffusion de contenus multimédia permet aussi d'anticiper des congestions, lorsque trop de terminaux sont associés à un seul point d'accès. Dans le cadre de la diffusion de contenus multimédia au sein d'une application dépendante du contexte, la prédiction de la mobilité permet également de diffuser le contenu approprié au contexte du terminal mobile.

4.3.2 Définition de seuils de préfetching

D. Charlet propose un mécanisme de préchargement des contenus multimédia [48]. Le préchargement intervient dans le contexte d'une architecture de caches répartis dans un réseau sans-fil. La problématique du préchargement est la gestion du handoff des clients mobiles vers lesquels les contenus multimédia sont diffusés. Les contenus considérés sont des flux hiérarchiques. Ce type de contenu permet de ne pas choisir une politique binaire qui consiste à prendre la décision de précharger totalement un contenu ou de ne rien précharger. Afin d'optimiser simultanément la qualité du service pour les clients et l'utilisation du réseau, deux seuils sont définis.

SysMoVie est une architecture de caches distribués dédiée à la diffusion de vidéo sur réseaux mobiles à grande échelle. SysMoVie intègre l'utilisation de la prédiction de la mobilité via des seuils de préchargement de flux hiérarchiques. En dessous du premier seuil, α , rien n'est préchargé dans le cache concerné. Entre les seuils α et T , seul le flux de base est préchargé. Au dessus du seuil T , tout le flux est préchargé. Il est possible de considérer plus de flux additionnels et plus de seuils, permettant de précharger une quantité progressive de données, fonction de la probabilité qu'a le terminal mobile de dépendre du cache concerné.

4.4 Modèles proposés

Nous avons choisi de nous baser sur les modèles de Markov. Dans un premier temps, nous présentons nos modèles basés sur les modèles de Markov sans prise en compte du temps de présence dans un état. Nos modèles sont k -past, k -to-1-past et k -to-1-past*. Dans un second temps, nous décrivons les améliorations de nos modèles pour la prise en compte du temps.

4.4.1 Modèles de Markov et mobilité, sans prise en compte de la temporalité

Nous nous basons sur les modèles de Markov pour établir la prédiction de la mobilité. Dans notre contexte, nous fixons un paramètre N qui est l'ordre le plus haut pour lequel nous construi-

sons les modèles de Markov. Les modèles sont construits pour tous les ordres de 1 à N . Pour construire les modèles de Markov, nous parcourons les fichiers de journalisation dans lesquels figurent (au moins) les données suivantes :

- ▶ la date d'arrivée dans un état (date unix) ;
- ▶ le lieu d'arrivée, c'est-à-dire l'état ;
- ▶ si les données sont relatives à plusieurs utilisateurs, l'utilisateur courant doit également figurer dans les données.

```
(...)
1206862437;Epsilon
1206971850;Hall
1206971880;Couloir RdC
1206971900;Couloir étage
1206971930;Bureau 1100
1206975521;Couloir étage
1206975547;Couloir RdC
1206975555;Hall
1206975567;Epsilon
(...)
```

FIG. 4.1 – Exemple de données de journalisation.

L'état ϵ , correspondant à l'absence du terminal mobile dans le système de gestion de la mobilité est l'état d'entrée et de sortie du modèle de Markov. Quand une occurrence de ϵ est rencontrée, les derniers états construits sont ajoutés et toutes les données temporaires sont purgées. La figure 4.1 contient un exemple fictif de données journalisées ne concernant qu'un utilisateur.

4.4.2 Modèle k-past : historique de longueur constante

Nous basons la prédiction de la mobilité sur 3 modèles dérivés des modèles de Markov. Le modèle **k-past** se base sur un modèle de Markov d'ordre k auquel un seuil σ , compris entre 0 et 1, est adjoint. Les modèles de Markov sont construits par apprentissage des données enregistrées lors de la journalisation.

Définition et construction

Nous définissons Φ égal à l'ensemble des états possibles dans l'espace physique. L'espace physique est la réalité. Par opposition, E_k est l'ensemble des états du modèle de Markov d'ordre k . Le cardinal de Φ est $\Omega(E_k) \leq \phi^k$. L'état ϵ modélise la sortie du terminal mobile hors du système. Il correspond à la perte de la trace du terminal mobile par le système de géolocalisation. Il correspond également à l'absence du terminal mobile avant son entrée dans le système et est préfixé automatiquement à une trajectoire lors de la détection du terminal mobile.

Le modèle k -past est défini par $(E_k, T_k, I_k, \Pi_k, F_k, \sigma)$. E_k est l'ensemble des états du modèle. Les états du modèle sont formés comme dans un modèle de Markov d'ordre k : leur étiquette est un ensemble ordonné de k états physiques. T_k est la matrice de transitions entre les états du modèle. Les transitions sont étiquetées par leur probabilité. I_k est l'ensemble des états initiaux du modèle. Π_k est la distribution de probabilité des états initiaux du système. F_k est l'ensemble des états terminaux. Notons que, dans le modèle 1-past, $E_1 \subset \Phi$, du fait de la nature des états d'ordre 1. Également, dans le modèle de Markov d'ordre 1, $I_1 = F_1 = \{\epsilon\}$. Pour tout $k \geq 1$, les étiquettes des états initiaux du modèle commencent par ϵ et les étiquettes états finaux du modèle finissent par ϵ .

La construction d'un modèle de Markov d'ordre k fait appel à la constitution de n -uplets où $n = k$. Les n -uplets sont formés de toutes les successions de n états journalisés, ordonnés et ne contenant pas l'état ϵ excepté en première ou en dernière position. Suivant l'exemple de données journalisées de la figure 4.1, les triplets possibles sont :

- $(\epsilon; \text{Hall}; \text{Couloir Rdc})$;
- $(\text{Hall}; \text{Couloir Rdc}; \text{Couloir étage})$;
- $(\text{Couloir Rdc}; \text{Couloir étage}; \text{Bureau 1100})$;
- $(\text{Couloir étage}; \text{Bureau 1100}; \text{Couloir étage})$;
- $(\text{Bureau 1100}; \text{Couloir étage}; \text{Couloir RdC})$;
- $(\text{Couloir étage}; \text{Couloir RdC}; \text{Hall})$;
- $(\text{Couloir RdC}; \text{Hall}; \epsilon)$.

Les n -uplets servent de base dans l'étude des transitions afin de les représenter dans des modèles de Markov. Un modèle k -past est construit à partir de l'ensemble des n -uplets. L'algorithme considère les données d'un seul utilisateur. On construit d'abord la matrice des occurrences O_k , qui porte les valeurs $o_{k_{i,j}}$ égales aux occurrences observées de la transition de l'état d'indice i à l'état d'indice j .

CONSTRUIT_MATRICE_OCCURRENCES($E_k, \text{liste_k_uplets}, k$)

```

1   $O_k$  : matrice carrée d'entiers
2   $source, destination$  :  $k$ -uplets
3   $i, j, k$  : entier
4   $destination \leftarrow \text{liste\_k\_uplets}[0]$ 
5  for  $i \leftarrow 1$  to  $\text{longueur}(\text{liste\_k\_uplets})$ 
6  do  $source \leftarrow destination$ 
7      $destination \leftarrow \text{liste\_k\_uplets}[i]$ 
8     if  $destination[0] \neq \epsilon$ 
9         then  $j \leftarrow \text{indice\_etat}(E_k, source)$ 
10          $k \leftarrow \text{indice\_etat}(E_k, destination)$ 
11          $O_k[j][k] \leftarrow O_k[j][k] + 1$ 
12 return  $O_k$ 
```

FIG. 4.2 – Construction du modèle de Markov d'ordre K .

L'algorithme 4.2 décrit la construction de la matrice O_k . L'algorithme reçoit la liste des états du modèle, c'est-à-dire une liste ordonnée des k -uplets, la liste des k -uplets et le paramètre k . La matrice O_k est carrée, de dimension égale au nombre d'états et initialisée à zéro. Les variables

source et *destination* définissent l'état source et l'état destination d'une transition. Si $k > 1$, ces deux variables sont composées de plusieurs éléments de l'ensemble Φ . L'algorithme ajoute une occurrence à la transition identifiée par les positions des états *source* et *destination* dans la liste d'états. L'ajout n'est pas effectué si la destination commence par ϵ car il s'agit d'un début de pérégrination.

À partir de la matrice des occurrences O_k , on peut calculer la matrice des probabilités de transitions T_k . Pour la calculer, on applique à chaque élément de la matrice O_k la formule suivante :

$$t_{k_{i,j}} = \frac{o_{k_{i,j}}}{\sum_{m=1}^{\Omega(E_k)} o_{k_{i,m}}}$$

Application à la prédiction de la mobilité

```

K-PAST( $E_k, T_k, \text{etat\_observe}$ )
1  prediction : liste d'états physiques
2  destination  $\leftarrow$  liste_k_uplets[0]
3   $i, j$  : entier
4  prediction  $\leftarrow \emptyset$ 
5   $i \leftarrow \text{indice\_etat}(\text{etat\_observe})$ 
6  if  $i \neq -1$ 
7    then for  $j \leftarrow 0$  to (longueur( $E_k$ ) - 1)
8      do if  $T_k[i][j] \geq \sigma$ 
9        then destination  $\leftarrow \text{etat\_label}(E_k, j)$ 
10       ajoute(prediction, destination[ $k - 1$ ])
11 return prediction

```

FIG. 4.3 – Prédiction de la mobilité, modèle k-past.

La prédiction basée sur le modèle k -past suit l'algorithme 4.3. L'étape préliminaire à l'exécution de l'algorithme k -past est la constitution de l'historique des k dernières positions connues du terminal mobile. L'algorithme k -past prend en paramètres la liste ordonnée des états du modèle E_k , la matrice des transitions T_k et l'état observé, c'est-à-dire l'historique du terminal mobile, pour lequel on cherche à établir la prédiction. Le résultat de l'algorithme sera contenu dans la variable *prediction*. La première étape est la recherche de l'état du modèle égal à l'historique du terminal mobile. La recherche de l'état renvoie son indice dans T_k . Si l'état observé n'existe pas, la valeur -1 est renvoyée. Dans le cas où l'état existe, chaque élément de la ligne correspondante de la matrice T_k est comparé avec le seuil de probabilité σ . Si la valeur de l'élément est supérieure ou égale à σ , l'état correspondant est recherché et le dernier état physique de l'étiquette est extrait. Il s'agit d'une destination suffisamment probable. L'état est ajouté à la liste des réponses incluses dans la prédiction. Quand toute la ligne de la matrice a été parcourue, la variable *prediction* est renvoyée par le modèle. Si aucun état n'est assez probable (par rapport à σ) ou que l'historique du terminal n'a pas été trouvé dans E_k , un ensemble vide est retourné par le modèle.

Les états formant l'ensemble de prédiction sont les états pour lesquels une action de handoff sera effectuée. En particulier, dans le cas de la gestion de caches distribués [48], il s'agit du

préchargement des données multimédia dans les caches voisins où le terminal mobile a le plus de probabilités d'être.

4.4.3 k-to-1-past, k-to-1-past* : historiques de tailles variables

Le défaut d'un modèle k -past est d'ignorer des historiques de déplacement inférieurs à sont ordre. En effet, un ordre trop grand a potentiellement un grand nombre d'état, de l'ordre de $\Omega(\Phi)^k$. Par conséquent, à quantité égale, les données d'apprentissage couvriront moins d'états du modèle et il y aura moins d'occurrences de chaque état du modèle. Pour pallier cette propriété, nous proposons de combiner plusieurs ordres de k -past. Les modèles **k-to-1-past** et **k-to-1-past*** sont composés des modèles i -past, $1 \leq i \leq k$.

Dans un premier temps, nous présentons le modèle k -to-1-past, semblable à un modèle AKMM. Dans un second temps, nous présentons le modèle k -to-1-past*, plus fiable que le modèle k -to-1-past.

K-to-1-past

Le modèle k -to-1-past est au modèle k -past ce que le modèle AKMM est au KMM. Le modèle k -to-1-past contient des modèles i -past pour $i \in [1; k]$. Il utilise un seuil σ unique pour l'ensemble des modèles k -past.

```

K-TO-1-PAST(modeles, transitions, etat_observe,  $k_{max}$ )
1  prediction : liste d'états physiques
2  current_k : entier
3  prediction  $\leftarrow \emptyset$ 
4  current_k  $\leftarrow k_{max}$ 
5  while (current_k > 0) ET (prediction  $\leftarrow \emptyset$ )
6  do prediction  $\leftarrow kpast(modeles[current\_k - 1], transitions[current\_k - 1], etat\_observe)$ 
7     current_k  $\leftarrow current\_k - 1$ 
8  return prediction

```

FIG. 4.4 – Prédiction de la mobilité, modèle k-to-1-past.

L'algorithme 4.4 cherche des états dans les modèles d'ordres k à 1. Dès qu'un des modèles a pu retourner un résultat, l'algorithme s'arrête. Cela permet de ne pas renvoyer une prédiction vide alors que des modèles d'ordres inférieurs seraient susceptibles de prédire un ensemble d'états probables. Cependant, la prédiction renvoyée par le premier modèle qui en est capable n'est pas toujours juste, souvent parce que les états du modèle ne sont pas fréquents et, par conséquent, statistiquement peu fiables. C'est pourquoi, nous proposons l'algorithme k -to-1-past* pour pallier cette propriété.

K-to-1-past*

Le modèle k -to-1-past* est formé des modèles i -past avec $1 \leq i \leq k$. Comme le modèle k -to-1-past, il comporte un seuil σ . L'objectif du modèle k -to-1-past* est d'utiliser les prédictions émises par l'ensemble des modèles qui le composent.

K-TO-1-PAST(*modeles*, *transitions*, *etat_observe*, k_{max})

```

1  prediction : liste d'états physiques
2  current_k : entier
3  prediction  $\leftarrow \emptyset$ 
4  current_k  $\leftarrow k_{max}$ 
5  while current_k > 0
6  do prediction  $\leftarrow$  prediction  $\cup$  kpast(modeles[current_k - 1], transitions[current_k - 1], etat_observe)
7     current_k  $\leftarrow$  current_k - 1
8  return prediction
```

FIG. 4.5 – Prédiction de la mobilité, modèle k-to-1-past*.

L'algorithme 4.5 expose le fonctionnement du modèle *k-to-1-past**. On constate que ce dernier fonctionne de manière similaire à *k-to-1-past*, à la différence qu'il ne s'arrête pas quand un modèle a pu émettre une prédiction. Il continue jusqu'à avoir testé tous les modèles qui le composent. Les résultats renvoyés par chaque modèle sont ajoutés dans le résultat global. De cette façon, quel que soit le modèle le plus pertinent pour un historique observé, sa prédiction sera intégrée à la prédiction finale.

4.4.4 Prise en compte de la date et des délais

Les modèles *k-past*, *k-to-1-past* et *k-to-1-past** permettent d'émettre une prédiction basée sur l'historique des déplacements d'un terminal mobile. Cependant, d'autres facteurs peuvent et doivent être pris en compte. En particulier, l'heure à laquelle l'historique des déplacements est observé est importante car de nombreux événements rythment la vie de l'utilisateur d'un terminal mobile et vont influencer sur ses déplacements. De plus, le délai à l'issue duquel le terminal mobile réalisera la transition prédite est critique également car une politique de handoff appliquée trop tôt est inutile, tout comme elle sera obsolète si elle est appliquée trop tard.

C'est pourquoi, en nous basant sur le modèle bayésien de J. Petzold et al., nous enrichissons le modèle *k-to-1-past** des dates de présence dans les états ainsi que du délai des transitions. Le modèle résultant est le **modèle de prédiction de la mobilité *k-to-1-past** temporel**.

Intégration des dates

Nous choisissons d'utiliser les jours de la semaine et les heures de la journée. Les étiquettes des états de chaque modèle sont enrichies du jour de la semaine, compris entre 0 (lundi) et 6 (dimanche) et de l'heure du jour, comprise entre 0 et 23. En effet, non seulement l'heure détermine les actions de l'utilisateur du terminal mobile, mais le jour est également un paramètre important. Par exemple, un employé de bureau ira manger vers 12h en semaine, et n'ira pas en direction de son lieu de travail le week-end.

Le modèle se construit de manière similaire au modèle *k-past* basique. Il suffit de considérer le jour de la semaine comme $(k + 1)$ ème élément d'un état et l'heure du jour comme $(k + 2)$ ème élément de l'état. La prédiction se fait alors toujours suivant le même algorithme (cf. algorithme 4.3), dans lequel la recherche de l'indice de l'état dans la liste tiendra compte des paramètres "jour de la semaine" et "heure".

Intégration des délais

La prise en compte des délais de transition nécessite d'ajouter deux éléments au modèle k -past. Le premier, Δ , est la liste des délais pour aller d'un état à l'état suivant. Il est calculé pour chaque transition par soustraction de la date d'arrivée dans l'état source à la date d'arrivée dans l'état destination. L'ensemble de ces temps est enregistré dans la liste de délais Δ .

Le second paramètre du modèle est le seuil de confiance souhaité pour la prise en compte du délai. En effet, on peut déclencher immédiatement la politique de handoff vers les états déterminés par la prédiction. Cependant, une politique aussi directe va surcharger le réseau avec des anticipations de handoffs intempestifs alors que ceux-ci pourraient être déclenchés au dernier moment. Le seuil de confiance va définir quel sera le délai à partir duquel le handoff sera déclenché. Le délai est basé sur les délais enregistrés. Le paramètre de confiance est δ . Il permet de choisir la borne inférieure de l'effectif de Δ représentant $\delta\%$ de Δ . L'effectif est choisi dans la partie supérieure de l'ensemble Δ .

4.5 Validation des modèles de prédiction de la mobilité

Afin de valider nos propositions, nous les avons testées. Nous avons également testé certains travaux sur lesquels se basent nos contributions. Les tests ont été effectués sur deux ensembles de données. Nous n'avons pas fait appel à des simulations. Dans cette section, nous exposons premièrement pourquoi avoir choisi des données issues de vraies pérégrinations. Deuxièmement, nous présentons les données employées et leur structure. Troisièmement, nous énumérons les modèles implémentés et fournissons les résultats obtenus, suivis par leur analyse.

4.5.1 Données de cas réels et simulation

Nous privilégions des données réelles à des données issues de simulations. Bien qu'il existe des simulateurs dédiés à la mobilité, tels que GIZMO et GLOMOSIM par exemple, ceux-ci ont pour vocation de simuler l'impact de la mobilité sur les transmissions dans les réseaux sans-fil. Nous évitons l'emploi de tels modèles car ils suivent des modèles statistiques pour générer des pérégrinations. Le problème sous-jacent est que notre problématique dans la prédiction de la mobilité est de modéliser les pérégrinations et d'en tirer des données statistiques exploitables pour la prédiction. Or, utiliser des pérégrinations générées suivant un modèle connu causera un biais dans l'évaluation de la pertinence d'un modèle.

À l'inverse, des données réelles sont issues de l'étude des pérégrinations de personnes qui ne suivent pas une règle particulière parce qu'elles sont soumises à des désirs et des contraintes complexes. Même si une personne suit généralement une règle particulière, il y aura des déviations dans les observations habituelles. Ces déviations sont difficiles à générer avec un simulateur. Nous avons choisi pour cela de nous baser sur des données journalisées à partir d'expériences réelles.

Les deux jeux de données que nous utilisons dans nos tests sont nommés **benchmarks** et **NokiaContextData**. Tous deux proviennent du site de l'institut pour l'informatique pervasive¹². Les deux jeux de données contiennent les données et un fichier de description du contenu des fichiers. Les données **NokiaContextData** contiennent des informations relatives au statut

¹²http://www.pervasive.jku.at/Research/Context_Database/index.php

GSM ainsi que de divers capteurs, formant 3 thèmes de données. Les données temporelles sont les suivantes :

- le temps unix ;
- le jour de la semaine ;
- le jour Mois période ;
- l’heure ;
- la demie-heure.

Les données GSM sont :

- la cellule ;
- le code de zone.

Enfin, les données de capteurs sont :

- la température ;
- l’humidité ;
- le point de rosée ;
- la pression ;
- l’orientation ;
- l’activité ;
- le niveau sonore.

Les champs sont séparés par des tabulations, tous sont présents à chaque ligne des fichiers de données. Le premier bloc de données concerne la date d’observation de l’enregistrement. La période distingue s’il s’agit de la nuit, du matin, du midi ou de l’après-midi. La cellule et le code de zone forment une donnée de positionnement du terminal GSM. Enfin, les données additionnelles quantifient l’environnement du terminal. Nous ne nous sommes pas intéressés à ces dernières.

Le second jeu de données [50] concerne les déplacements du personnel d’un laboratoire de recherche. Quatre personnes ont indiqué en cliquant sur un plan affiché sur leur PDA la salle dans laquelle ils pénétraient au gré de leurs mouvements. À chaque entrée dans une salle, la personne, la date et le lieu indiqué sont écrits dans un fichier qui a la forme suivante :

```
yyyy.mm.dd hh:mm:ss;position;personne;date unix
```

yyyy.mm.dd est la date du déplacement, hh :mm :ss étant son heure. La position est le nom de salle dans laquelle la personne a signalé son entrée. La personne est définie par une lettre parmi A, B, C et D. La date unix est exprimée en millisecondes depuis le 1^{er} janvier 1970.

4.5.2 Expérimentations

Pour les expérimentations, nous testons les modèles que nous avons présentés avec les jeux de données de Nokia et d’Augsburg. Plusieurs séries de tests sont effectuées. La première série consiste à fournir 75% d’un fichier de données pour l’apprentissage du modèle. Les 25% restants sont ensuite fournis au modèle pour la validation. Une série de tests restreinte aux données d’Augsburg consiste à utiliser les données de l’automne pour l’apprentissage et les données de l’été pour la validation. La troisième série de tests teste les données au fur et à mesure de leur lecture. Dès qu’une donnée est testée, elle est ensuite intégrée au modèle. C’est un apprentissage continu, permettant d’ajuster le modèle en fonction des changements d’habitudes

des utilisateurs. La quatrième série de tests consiste à intégrer les données dans le modèle par groupe de 50 et à tester à chaque fois l'ensemble des données, y compris celles ayant participé à l'apprentissage.

Modèles de Markov atemporaux

Dans ces tests, la précision est le taux de prédictions exactes sur la totalité des transitions observées. Par conséquent, les modèles k -past d'ordre supérieur à 1 perdent $k - 1$ tentatives par pérégrination, le temps de constituer un historique complet. En revanche, les modèles k -to-1-past et k -to-1-past* fonctionnent dès la première observation, en n'utilisant que les premiers modèles i -past qui les composent. Bien que cette forme d'évaluation pénalise le modèle k -past, elle est logique car, dans la politique de handoff, si le système ne peut pas émettre de prédiction, il y a interruption des services.

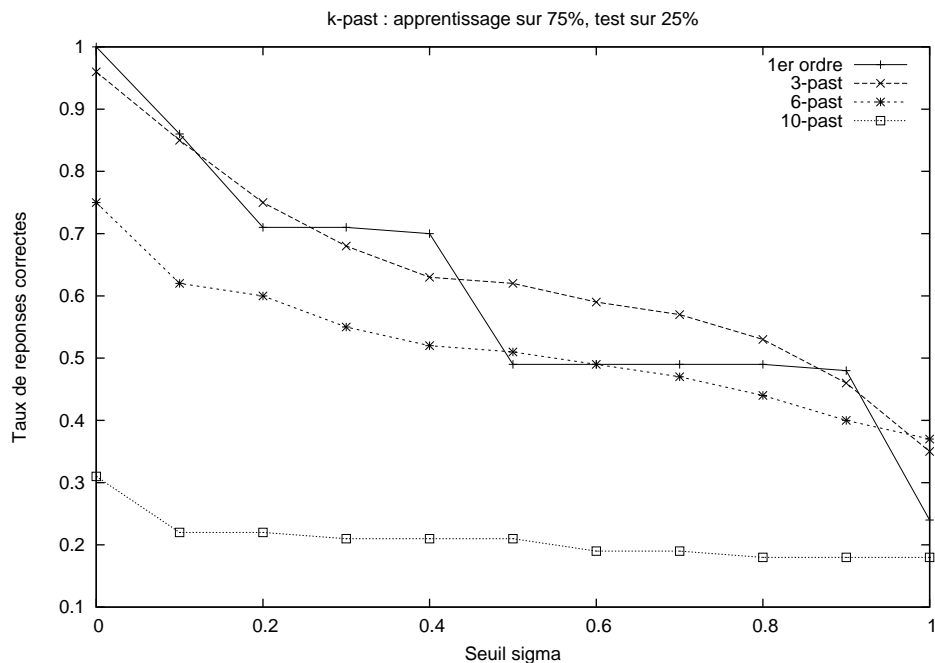


FIG. 4.6 – Modèle k -past, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.

La figure 4.6 expose les résultats de la prédiction de la mobilité par le modèle k -past. Le seuil de déclenchement σ apparaît en abscisse. L'axe des ordonnées donne la précision atteinte, c'est-à-dire le taux de prédictions correctes sur l'ensemble des transitions testées. On remarque que la précision n'augmente pas parallèlement à l'augmentation de l'ordre du modèle. Pour un modèle de Markov de troisième ordre, la précision est parfois supérieure, parfois inférieure à celle du modèle d'ordre 1. Le modèle d'ordre 6 quant à lui n'est plus précis que pour $\sigma = 0,5$ et $\sigma = 1,0$. Dans les autres cas, sa précision reste inférieure à celle des modèles d'ordres 1 et 3. On remarque également que le modèle d'ordre 10 a une précision très faible, systématiquement inférieure à celles des autres modèles. Ces observations s'expliquent en fonction de deux éléments. D'une

part, augmenter l'ordre du modèle conduit à augmenter la longueur des états. Par conséquent, alors qu'il suffit d'un état observé pour émettre une prédiction avec le modèle d'ordre 1, il en faut N pour le modèle d'ordre N . Par conséquent, pour le modèle d'ordre N , $N - 1$ transitions voient leur prédiction inexacte car absente. D'autre part, augmenter l'ordre du modèle augmente le nombre d'états et réduit d'autant leur fréquence d'apparition dans les données d'apprentissage et dans le test. Par conséquent, plus l'ordre augmente, plus il est difficile d'observer à nouveau un état du modèle dans les données de test. Hors, s'il est impossible de trouver l'état observé dans le modèle, aucune prédiction n'est effectuée donc le protocole de test considère que la prédiction est fausse.

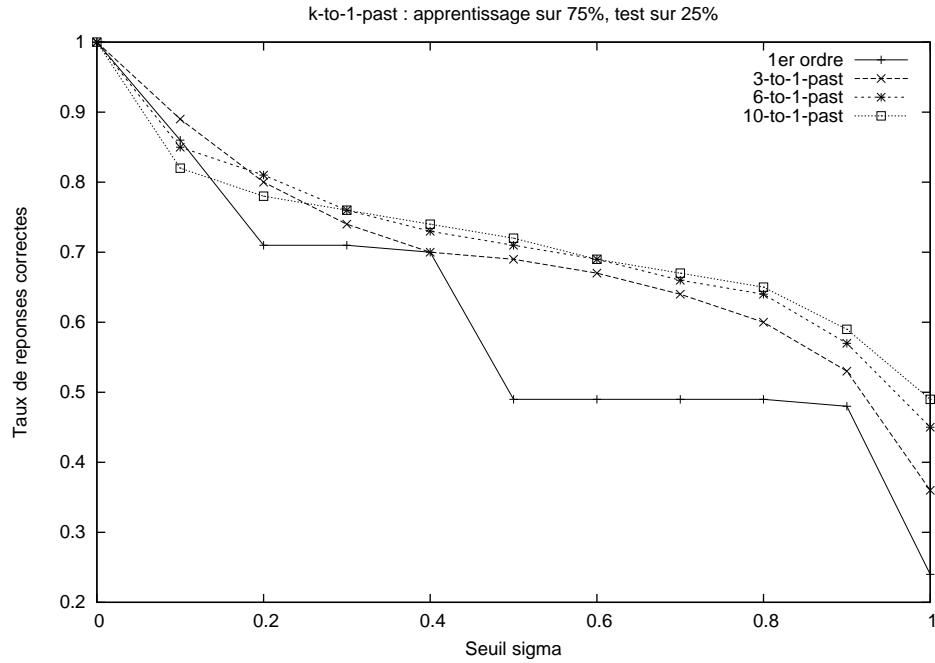


FIG. 4.7 – Modèle k-to-1-past, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.

La figure 4.7 expose les résultats de la prédiction de la mobilité par le modèle k-to-1-past. Concernant les résultats de ce modèle, on remarque premièrement que l'augmentation de l'ordre du modèle ne dégrade pas sa précision. Au contraire, le modèle 10-to-1-past est généralement plus précis que les modèles d'ordres inférieurs. En particulier, dès que $\sigma \geq 0,4$, le modèle d'ordre 10 est toujours plus précis. On peut expliquer ce résultat par rapport aux résultats du modèle k-past : comme le modèle recherche dans tous les modèles de Markov d'ordres inférieurs ou égal à lui-même, par ordre décroissant et jusqu'à trouver un résultat, les observations ont toujours une chance de mener à une prédiction, au minimum par le modèle d'ordre 1. De la même façon, si l'observation ne permet pas de trouver un état dans le modèle d'ordre correspondant, les modèles d'ordres inférieurs ont également une chance de contenir un état permettant d'émettre une prédiction. Par conséquent, les deux principaux inconvénients du modèle k-past sont palliés par le modèle k-to-1-past.

La figure 4.8 expose les résultats de la prédiction de la mobilité par le modèle k-to-1-past*.

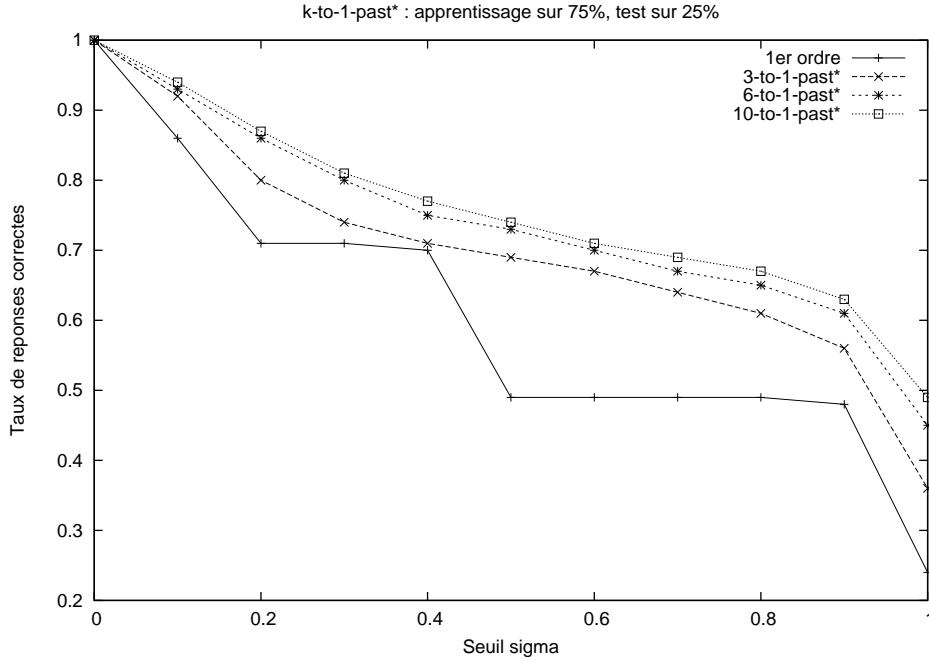


FIG. 4.8 – Modèle k-to-1-past*, précision de prédiction avec apprentissage sur 75% des pérégrinations, test sur 25%.

Une légère amélioration par rapport au modèle k-to-1-past est visible. En particulier, plus un modèle est d'ordre élevé, plus il est précis. En effet, l'agrégation des prédictions effectuées par les modèles de Markov d'ordres 1 à N pour un modèle N -to-1-past* conduit à contenir au moins tous les états choisis par le modèle k-to-1-past. Par conséquent, la précision d'un modèle N -to-1-past* est nécessairement supérieure ou égale à celle d'un modèle N -to-1-past.

La figure 4.9 décrit les résultats de la prédiction de la mobilité lorsque l'apprentissage des transitions est effectué à partir des profils *fall* et que le modèle résultant est testé avec les transitions des profils *summer*. La figure est composée de 3 sous-figures qui affichent les résultats des modèles k-past (fig. 4.9(a)), k-to-1-past (fig. 4.9(b)) et k-to-1-past* (fig. 4.9(c)). On remarque globalement sur toutes les courbes que la précision décroît lorsque la valeur de σ augmente. En effet, un état est sélectionné pour la prédiction si sa probabilité dans le modèle est égale ou supérieure à σ . Par conséquent, plus σ est élevé, plus il est difficile pour un état d'être sélectionné donc, plus la précision va diminuer. On remarque également sur toutes les courbes que les précisions des modèles de premier ordre, à contexte de test identique, sont égales, quel que soit le modèle observé. En effet, k-to-1-past utilise la première prédiction possible. À l'ordre 1, seul le modèle de Markov de premier ordre peut servir. Donc sa prédiction est identique à celle produite par le modèle 1-past. La même observation vaut également pour le modèle k-to-1-past* d'ordre 1.

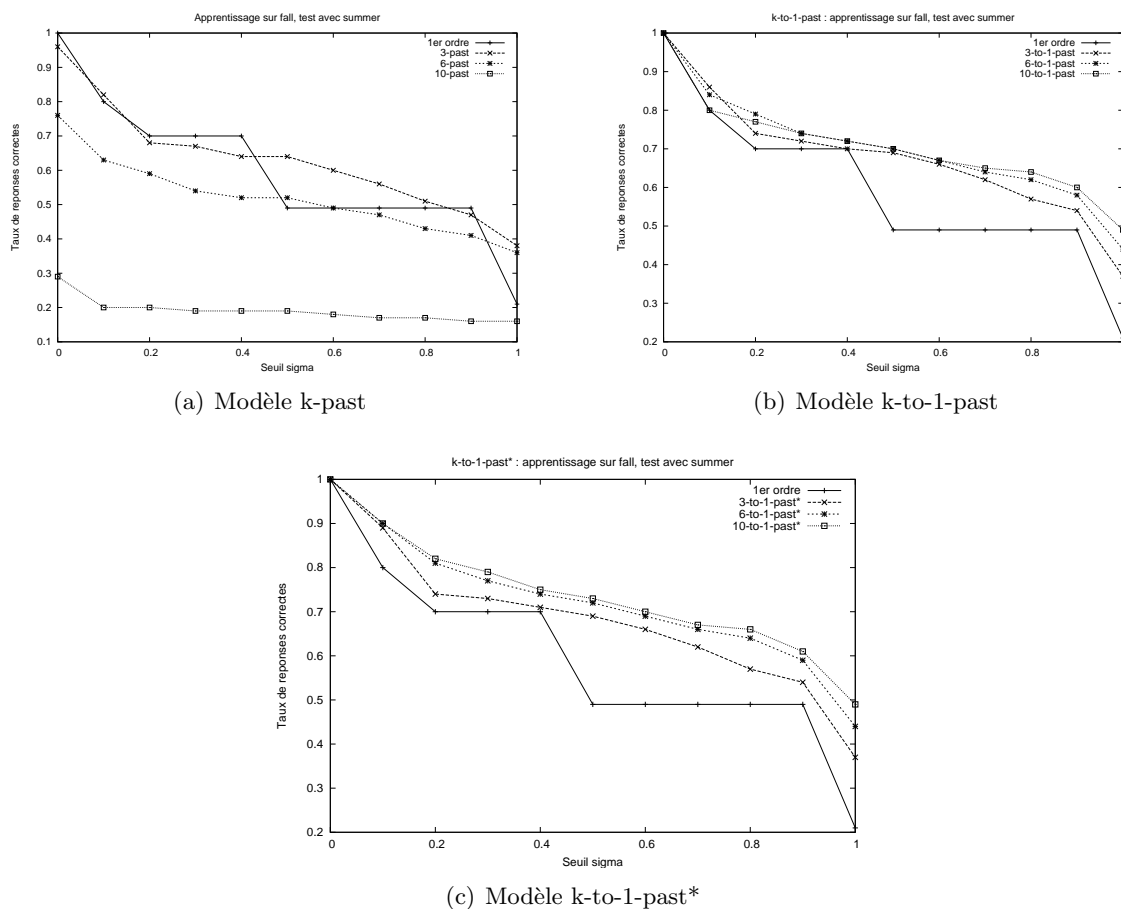


FIG. 4.9 – Tests des modèles atemporaux : apprentissage sur le profil fail, test sur le profil summer.

Réseau bayésien et prise en compte du temps

Dans les tests avec la prise en compte du temps, la précision est mesurée par le taux de bonnes prédictions pour lesquelles la politique de handoff a été déclenchée à temps. Le taux est calculé par rapport à l'ensemble des prédictions tentées, comme dans le test des modèles atemporaux.

La figure 4.10 expose les résultats respectifs des modèles k-past (sous-fig. 4.10(a)), k-to-1-past (sous-fig. 4.10(b)) et k-to-1-past* (sous-fig. 4.10(c)) sur les fichiers individuels. Pour chaque fichier, 75% des pérégrinations sont utilisées lors de l'apprentissage et les 25% restants sont confrontés au modèle qui en résulte. On remarque que la précision est légèrement inférieure à celle obtenue par les modèles sans prise en compte du délai. Cette observation est logique car la prise en compte du délai ne peut que générer des prédictions fausses (au sens de l'évaluation) supplémentaires. Le reste des observations est semblable à l'analyse des modèles de prédiction de la mobilité sans prise en compte du délai. En effet, basiquement, les modèles restent les mêmes qu'auparavant.

Concernant l'intégration des jours de la semaine et des heures de la journée au modèle, nous ne donnons pas de résultat. En effet, l'augmentation du nombre des états par les paramètres

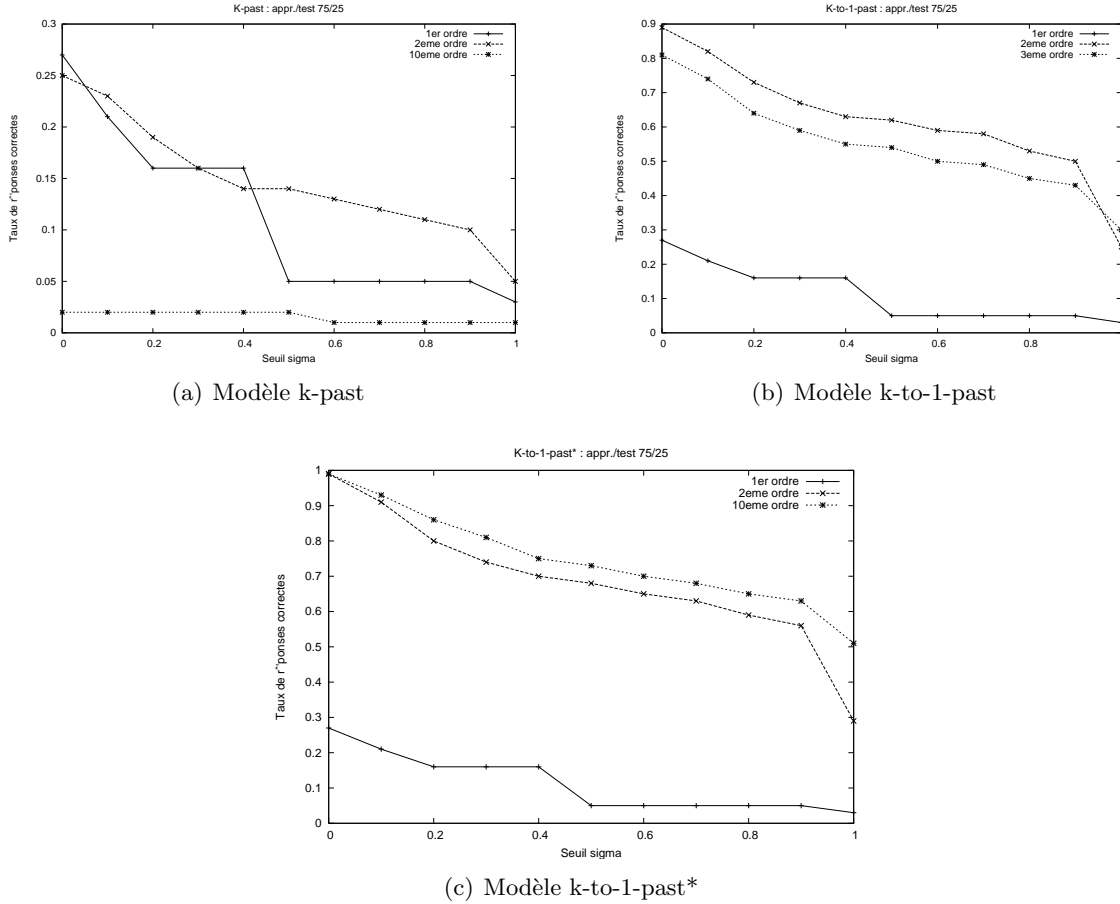


FIG. 4.10 – Précision avec prise en compte du délai (75/25).

temporels requis conduit à des modèles dans lesquels chaque état n'est présent qu'une fois et donc un état présent n'a qu'un successeur de probabilité égale à 1. On observe donc des résultats identiques quel que soit le seuil σ . De plus, les précisions sont très faibles, avec un maximum de 20% pour le modèle k-to-1-past*. Pour mieux évaluer et analyser un modèle prenant en compte les données temporelles, il serait nécessaire de disposer de données en quantités plus importantes.

Outre que les observations concernant les modèles atemporaux restent valables, on remarque que, globalement, les prédictions résultantes sont moins précises. En effet, il existe deux types d'erreurs : une mauvaise prédiction de l'état suivant, et une décision prise trop tard. À l'opposé, les modèles atemporaux ne peuvent commettre que la première erreur. De plus, l'intégration du temps dans les états conduit à produire encore plus d'états dans le modèle. On observe cependant que plus de 75% des handoff ont lieu dans de bonnes conditions quand on choisit une valeur de σ moyenne (0,5) avec les modèles prenant en compte le temps de passage entre les états.

Conclusion et perspectives

Dans ce chapitre, nous avons exposé des modèles de prédiction de la mobilité, basés sur les modèles de Markov. Les modèles que nous proposons doivent permettre de produire une prédic-

tion de la mobilité des terminaux mobiles. La prédiction doit être assez rapide pour s'effectuer avant que la transition réelle soit déclenchée. Le modèle k -past considère un historique de taille unique égale à k . Ce modèle n'est pas très précis utilisé seul, mais en combinant plusieurs degrés de modèle k -past, on obtient le modèle k -to-1-past, plus précis.

Le modèle k -to-1-past a pour défaut de parfois émettre une prédiction erronée. C'est pourquoi nous proposons le modèle k -to-1-past*, agrégeant les résultats de tous les k -past sous-jacents. Ce dernier modèle permet d'obtenir une meilleure précision que k -to-1-past et k -past, ainsi qu'une meilleure précision que les travaux de l'état de l'art.

Le modèle k -to-1-past* ne tenant pas compte du temps, nous l'avons enrichi afin de pallier ce manquement. Le modèle de prédiction de la mobilité k -to-1-past* temporel. Ce modèle atteint une précision suffisante (75% avec des paramètres bien choisis) pour envisager un déploiement dans un réseau sans-fil de diffusion de contenus multimédia riches. Grâce à ce modèle, 75% des déconnexions potentielles dues aux déplacements des terminaux mobiles seront traitées et évitées.

Nous prévoyons d'étudier l'encombrement du réseau par les procédures de *handover* en fonction des modèles. En effet, plus on sélectionne d'états pour y appliquer la prédiction de la mobilité, plus on a de chances de prendre la bonne décision. Cependant, on augmente en conséquence l'occupation du réseau par les communications entre les éléments du réseau et les éventuels transferts de cache.

De plus, dans le cadre de la diffusion de flux hiérarchiques dans un système de caches distribués, nous pouvons utiliser plusieurs seuils, chacun déterminant une qualité de flux. Dans un premier temps, nous nous intéresserons à l'usage de deux seuils, tels que décrits par D. Charlet dans sa thèse [47]. Par la suite, nous pouvons considérer autant de seuils qu'il n'y a de niveaux hiérarchiques dans la vidéo.

Enfin, les modèles construits à partir des pérégrinations des terminaux mobiles peuvent être réemployés dans la géolocalisation afin de remplacer la connexité des points de référence par le plan de déplacement usuel des terminaux mobiles. De cette façon, la métrique évolue, devenant une probabilité. Il est alors envisageable de construire un modèle de Markov à espace d'états cachés qui permette de réaliser la géolocalisation et la prédiction de la mobilité au sein d'un unique modèle.

Troisième partie

Mise en œuvre

Chapitre 5

Mise en œuvre du système de géolocalisation

Introduction

Dans ce chapitre, nous présentons les travaux d’implantation des contributions, en particulier concernant la géolocalisation de terminaux mobiles dans un réseau Wi-Fi. En effet, le modèle de géolocalisation FRBHM permet la localisation des terminaux mobiles avec suffisamment de précision pour envisager son utilisation dans des applications conscientes du contexte.

Le reste du chapitre est divisé en trois parties. La première partie présente la géolocalisation dans le contexte d’une plate-forme de diffusion de contenus multimédia vers des terminaux mobiles à grande échelle. La seconde partie décrit un projet de guide numérique mobile, applicable aux visites musées. La troisième définit les choix techniques pour l’implantation de la géolocalisation et de la prédiction de la mobilité dans les deux cas exposés.

5.1 GeoMoVie

GeoMoVie est un module de MoVie. MoVie est une plate-forme de diffusion de contenus multimédia, à partir de serveurs de vidéo, vers des terminaux mobiles. Cette plate-forme vise la diffusion à grande échelle, c’est-à-dire vers des milliers de terminaux mobiles. Afin de ne pas faire crouler les serveurs sous les requêtes, MoVie doit inclure un système de gestion de caches vidéo. De plus, devant l’hétérogénéité des terminaux mobiles (ordinateurs, téléphones, PDAs), les contenus diffusés doivent s’adapter au terminal client.

Pour cette raison, MoVie est décomposé en plusieurs modules :

- ▶ NetMoVie [51], un module de transcodage, visant à adapter les contenus aux terminaux ;
- ▶ SysMoVie [47], un module de gestion des caches vidéo, responsable également des politiques de handoff inter-caches ;
- ▶ WebMoVie, une interface Web qui permet aux terminaux clients de sélectionner leurs vidéos ;
- ▶ GeoMoVie, un module additionnel qui permet de connaître et anticiper la position des terminaux mobiles. GeoMoVie communique avec NetMoVie et SysMoVie pour leur fournir

des données de localisation.

Nous présentons dans un premier temps le contexte de MoVie. Dans un second temps, nous définissons rapidement le rôle de NetMoVie et sa relation à la position des terminaux mobiles. Dans un troisième temps, nous faisons de même pour le module SysMoVie. Enfin, nous présentons GeoMoVie, le module contenant nos contributions en termes de géolocalisation des terminaux mobiles.

5.1.1 Contexte

MoVie vise à diffuser des contenus multimédia riches, en particulier des vidéos et du son, vers des terminaux mobiles dans un réseau Wi-Fi. Les terminaux mobiles sont de types variés, tels que des assistants numériques personnels (PDA), des téléphones mobiles et des ordinateurs. Chacun de ces types d'appareils a ses propres caractéristiques en termes de puissance, connectivité, mémoire, codecs multimédia et résolution d'affichage. C'est pourquoi les contenus diffusés doivent être adaptés au type de client ciblé.

Par ailleurs, la qualité de la connexion entre un point d'accès du réseau et un terminal mobile varie au cours du temps. Pour limiter les saccades dans la lecture des contenus vidéo, un encodage hiérarchique est utilisé. Celui-ci est composé d'un flux de base, de qualité médiocre ainsi que d'au moins un flux d'amélioration. Chaque flux d'amélioration permet d'accroître la qualité de la vidéo. Parallèlement, l'occupation réseau est également augmentée. L'encodage hiérarchique a été choisi car il permet de pallier les baisses de débits, en particulier lors des congestions réseau, et de toujours envoyer au moins le flux de base. En effet, il vaut mieux envoyer une vidéo de qualité médiocre que rien du tout.

5.1.2 NetMoVie

NetMoVie permet d'adapter la vidéo au terminal mobile. NetMoVie repose sur un programme nommé **mixeur**, qui réceptionne une vidéo, la réencode pour le terminal mobile et envoie la vidéo résultante au client qui l'a requise. Le réencodage dépend du type de client et de codec lu par ce client. Il dépend également du débit dont le client peut disposer. En particulier, des baisses de débits liées au Wi-Fi apparaissent souvent car le Wi-Fi a une bande passante partagée entre tous les appareils.

Par conséquent, NetMoVie peut exploiter des données de localisation, en particulier si l'on peut prédire l'association de plusieurs terminaux au même point d'accès. Prédire la présence des terminaux mobiles permet d'anticiper les dégradations de débit disponible et d'adapter en conséquence la qualité diffusée aux terminaux mobiles.

5.1.3 SysMoVie

SysMoVie a pour objectif la gestion des caches vidéo. En effet, pour que des serveurs vidéo puissent servir des milliers de clients, il est nécessaire que leur contenu soit rapproché des points d'accès et des terminaux mobiles. C'est le rôle des caches vidéo de MoVie. La problématique pour les caches est liée à la mobilité des terminaux. En effet, au cours de leurs pérégrinations, les terminaux mobiles vont quitter la zone de couverture d'un cache pour passer à la zone de

couverture d'un autre cache. Ce phénomène, qui est le *handover*, doit être préparé pour éviter des interruptions de la lecture sur le terminal mobile.

La politique de handoff offerte par SysMoVie consiste à déterminer quelle qualité de vidéo sera préchargée dans les caches voisins du cache courant. La décision est dépendante de la probabilité de présence du terminal mobile dans chaque zone de couverture des caches voisins. La connaissance des probabilités de présence du terminal mobile est obtenue par l'étude de la mobilité des terminaux mobiles. Cette information peut s'obtenir grâce à un modèle de prédiction de la mobilité, tel que *k-to-1-past**.

5.1.4 GeoMoVie

GeoMoVie est le module de gestion de la mobilité des terminaux mobiles. Il est en charge de la localisation des terminaux et de la prédiction de leurs déplacements futurs. GeoMoVie dispose d'interfaces avec NetMoVie et SysMoVie, mais peut s'adapter à d'autres applications, à la manière d'un intergiciel.

GeoMoVie a plusieurs fonctions importantes. D'abord, il permet de géolocaliser les terminaux mobiles. La connaissance de la position d'un terminal mobile peut être transmise à divers acteurs du réseau, permettant d'adapter les services à la position du terminal mobile. La deuxième fonction importante de GeoMoVie est la journalisation des positions successives des terminaux mobiles. Ces données peuvent servir à établir des statistiques concernant les déplacements des terminaux mobiles. Elles permettent également de remplir le troisième rôle de GeoMoVie, l'apprentissage des données journalisées pour en tirer un modèle des déplacements des terminaux mobiles. Enfin, le modèle obtenu permet à GeoMoVie de remplir le rôle de prédiction de la mobilité.

5.2 GuiNuMo

Le projet GuiNuMo [30] consiste en l'élaboration d'un GUIde NUmerique MObile permettant de diffuser des informations multimédia géolocalisées et en mobilité. Il s'agit un guide électronique, principalement dédié à la visite d'espaces scénarisés tels que des expositions, des musées ou des parcs. Contrairement à ce que laisse sous entendre le terme "guide", GuiNuMo est conçu comme un assistant à la visite. Il est non directif et vise à ne pas accaparer l'attention de l'utilisateur. Il s'apparente plutôt à un conseiller disposant d'une mine d'information à propos de l'espace parcouru. L'utilisateur est libre de ses mouvements et le cheminement n'est pas imposé. GuiNuMo se présente sous la forme d'un assistant personnel (PDA) ou d'un "tablet PC ultra portable" dont l'une des principales caractéristiques est de disposer d'une interface d'accès à un réseau sans-fil.

La principale fonctionnalité de GuiNuMo est de s'adapter au contexte en s'appuyant sur la géolocalisation des usagers. Grâce à cette information la liste des média pouvant être acheminés à l'utilisateur est déterminée. Cette liste est constituée de média divers tel que du texte, des images fixes, des fichiers sonores ou vidéo. En fonctionnement standard, l'arrivée dans un lieu déclenche la lecture d'un fichier vidéo décrivant le contexte. L'utilisateur est ensuite libre de consulter les différents média qui lui sont proposés. La consultation d'un média peut entraîner de la part de GuiNuMo la proposition d'un conseil tel que "pour compléter votre visite vous pouvez vous

rendre à tel endroit” où vous pouvez consulter tel autre média. Ces conseils sont le fruit d’une scénarisation du lieu préalable au déploiement du guide numérique.

Dans la suite nous présentons les différentes phases de déploiement du guide numérique en nous appuyant sur une expérience de mise en place au sein du Musée de l’Aventure Peugeot à Montbéliard. Nous expliquons également les différentes interactions possibles entre le guide et l’usager. Nous décrivons également comment la mise en place du guide peut être utilisée pour analyser et éventuellement faire évoluer l’espace à visiter.

Le déploiement de GuiNuMo

Le déploiement de GuiNuMo au sein d’un espace à visiter s’élabore en plusieurs phases. Il tout d’abord nécessaire de mettre en place l’infrastructure de diffusion de l’information et de localisation des dispositifs mobiles. Il faut, concurremment, produire des média illustrant l’espace et définir un scénario associant les média avec les lieux ou les objets.

L’infrastructure de diffusion et de géolocalisation

GuiNuMo repose, tant pour la diffusion que pour la géolocalisation, sur un réseau Wi-Fi. Le principe du déploiement consiste donc à positionner au sein de l’espace visité un ensemble de points d’accès Wi-Fi. Le nombre et la position des bornes doivent être déterminés pour permettre une couverture optimale de l’espace en terme de diffusion, mais également pour permettre une localisation précise. Si le premier objectif est relativement simple à satisfaire, c’est-à-dire avoir une couverture complète du lieu en terme de diffusion, le deuxième objectif est plus complexe. En effet, permettre un positionnement pertinent au sein d’un espace complexe est une tâche ardue.

La première étape consiste en la mise en place de points d’accès Wi-Fi pour former un maillage triangulaire plus ou moins régulier, selon la possibilité. Une fois le maillage mis en place, il faut réaliser une phase de cartographie du lieu en relevant en des points, dont les coordonnées sont connues, les puissances des signaux de chacun des points d’accès afin de s’assurer d’une part de la couverture et d’autre part de la possibilité de pouvoir géolocaliser. Après avoir éventuellement fait évoluer le positionnement des bornes, une cartographie des puissances des signaux est stockée sur le serveur de géolocalisation.

Cette cartographie, associé à un mécanisme de localisation basée sur la trilatération dynamique des dispositifs mobiles permettra la localisation précise des visiteurs. La localisation sera d’autant plus fine que la taille des mailles est fine et que le pas de cartographie est fin.

la scénarisation

La scénarisation d’un lieu pour la mise en place d’un guide numérique peut être décomposée en trois phases. La première phase consiste en l’inventaire et la production des média devant être diffusés. La deuxième phase est la définition des zones et des objets devant faire l’objet d’une présentation par le guide. La troisième phase est la mise en relation des zones et des objets avec les média ainsi que la mise en relation des média entre eux et des lieux et des objets entre eux.

Une grande partie de la qualité de l’assistance fournie par le guide numérique repose sur deux piliers : d’une part la qualité de la localisation et d’autre part sur la capacité du guide

à proposer des parcours inédits ou des approfondissements de tel ou tel sujet. Pour être en mesure de faire ces recommandations le guide repose sur les liens proposés dans le scénario. Le principe est que, après une interaction de l'utilisateur tel que la visualisation d'une séquence vidéo le dispositif pourra proposer soit une autre séquence, soit un autre lieu ou objet à visiter sur un thème proche. Cette étape de scénarisation, même si elle faite en collaboration avec l'équipe mettant en place le dispositif, est principalement la tâche des muséologues.

5.3 Architecture

Dans cette section, nous présentons l'architecture mise en place pour assurer le fonctionnement d'un intergiciel de gestion de la mobilité. Dans un premier temps, nous exposons les réflexions nous ayant amenés au choix d'un intergiciel. Dans un second temps, nous décrivons l'architecture matérielle requise pour le fonctionnement des services de gestion de la mobilité. Dans un troisième temps, le cœur des travaux, c'est-à-dire l'infrastructure logicielle, est décrit et replacé dans le contexte des contributions.

5.3.1 Choix d'une approche orientée service

Ainsi que nous l'avons présenté dans les deux premières sections de ce chapitre, de nombreux services reposent ou peuvent reposer sur la gestion de la mobilité. La gestion de la mobilité étant complexe, il est souhaitable, en particulier pour les développeurs ne souhaitant pas y consacrer du temps, de disposer d'une "boîte noire". Cette boîte noire doit s'interfacer rapidement et facilement avec n'importe quelle application qui y aurait recours. Elle doit donc être interopérable avec toute application. C'est pourquoi nous choisissons une approche orientée service.

Nous choisissons d'intégrer le service de gestion de la mobilité à un intergiciel afin de faciliter son intégration à des applications et de simplifier son utilisation. Le service de gestion de la mobilité installé fonctionne en fournissant des données de positionnement et de prédiction de la mobilité aux applications qui y font appel. Les données sont données sous une forme simple et exploitable quel que soit l'application qui les requiert.

5.3.2 Architecture matérielle

L'architecture matérielle est un support à la mise en œuvre du service de gestion de la mobilité. La figure 5.1 décrit l'ensemble de l'architecture matérielle du service de gestion de la mobilité. On y trouve la partie infrastructure et la partie mobile. La partie infrastructure est composée d'un réseau câblé, de type fast ethernet, d'une machine serveur de géolocalisation et de points d'accès. Les points d'accès sont l'interface entre l'infrastructure du réseau et la partie mobile. La partie mobile du système est composée d'une flotte de terminaux mobiles. Ces derniers sont des ordinateurs portables, des téléphones mobiles, des assistants numériques personnels ou encore des consoles portables dotées d'une connectivité sans-fil.

La variété des terminaux mobiles dans un environnement ouvert, c'est-à-dire où les utilisateurs apportent leur propre terminal, donne une large importance au rôle de l'infrastructure. En effet, c'est sur cette dernière que les administrateurs du service auront le plus de possibilités de modification. Par ailleurs, certains déploiements varient légèrement de ce cas de figure. Il s'agit

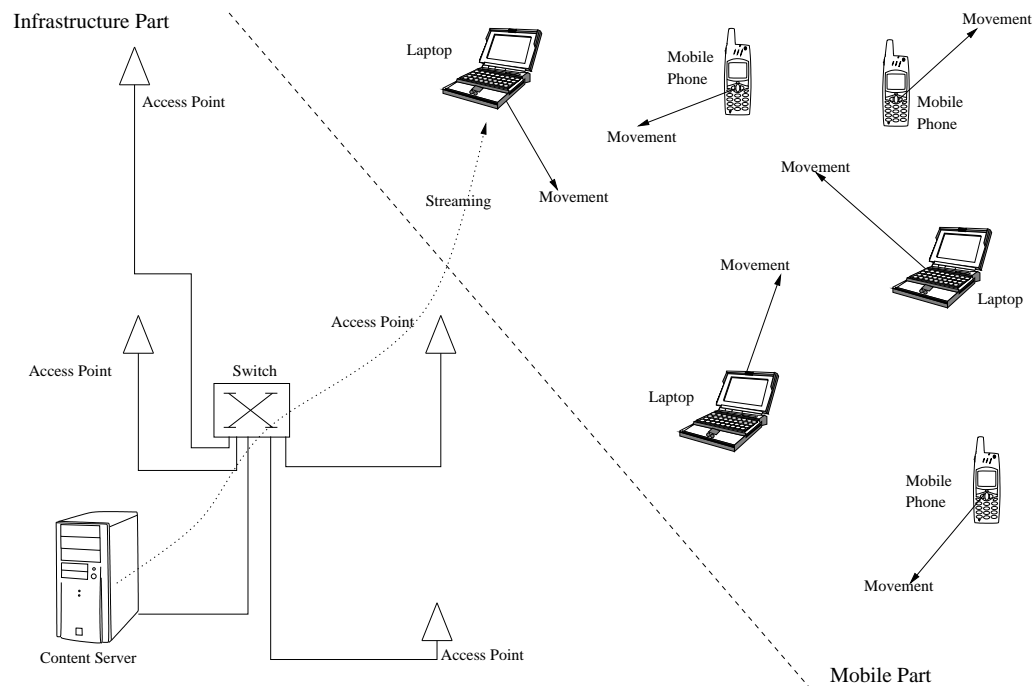


FIG. 5.1 – Architecture matérielle du service de gestion de la mobilité.

notamment des visites assistées, dans lesquelles le musée fournit également les terminaux. Dans ce cas, les administrateurs ont également le pouvoir de modifier les terminaux. Ces types de déploiements vont influencer sur l'infrastructure logicielle.

5.3.3 Architecture logicielle

L'architecture logicielle est décomposée en plusieurs modules. En effet, les différentes formes d'infrastructures matérielles vont conduire à une diversification des modules logiciels. Le premier module que nous allons présenter est le serveur de géolocalisation. C'est le cœur du système, dans lequel sont implémentées les méthodes de géolocalisation. Nous présentons ensuite un ensemble de modules dédiés à la mesure de la puissance des signaux reçus. Plusieurs modules sont requis, en fonction de l'infrastructure matérielle et de la position de la mesure, soit sur les terminaux, soit sur l'infrastructure.

Actuellement, seul le système de géolocalisation est implanté. Des programmes basiques de test des algorithmes de prédiction de la mobilité existent, mais ils ne sont pas encore intégrés au service de gestion de la mobilité.

Le serveur de géolocalisation

Le serveur de géolocalisation est la pièce maîtresse du service de gestion de la mobilité. Il est actuellement écrit en C++. C'est le serveur qui permet de détecter la présence et la position des terminaux mobiles. Il comporte plusieurs méthodes de localisation. En particulier, le FBCM et le FRBHM sont intégrés au serveur. Le serveur est un processus multi-threads. Il est en

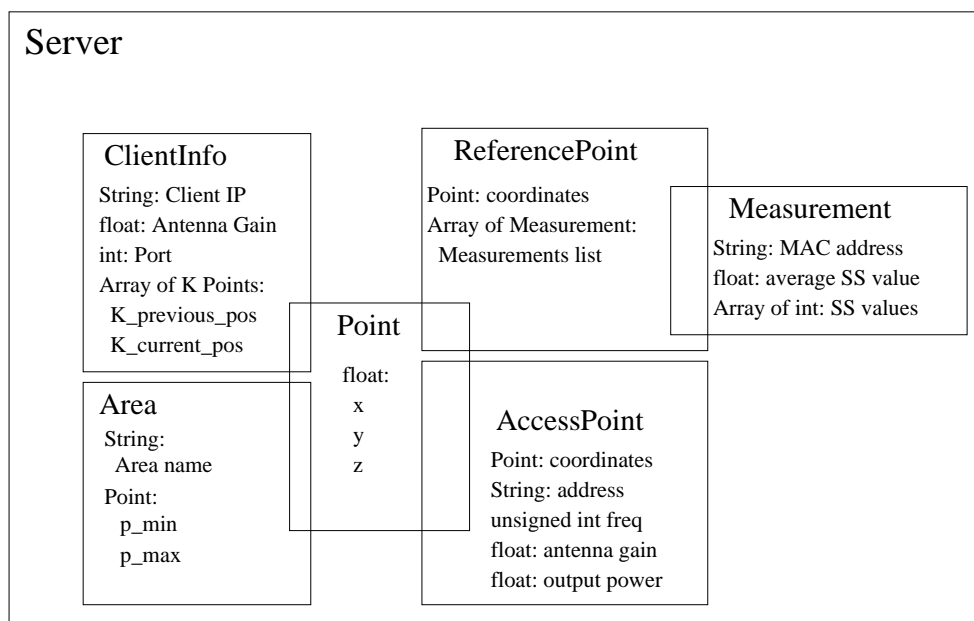


FIG. 5.2 – Architecture fonctionnelle de la plate-forme de test : classes utilisées.

attente des connexions des terminaux mobiles. La figure 5.2 indique les classes présentes dans le serveur de géolocalisation. Les informations nécessaires de chaque terminal mobile sont stockées par le serveur. C'est le rôle de la classe **ClientInfo**. Un client est défini par un historique fixe, nécessaire à l'application des algorithmes de Viterbi. Il est également défini par une adresse IP et un port sur lesquels le serveur pourra les contacter. Le gain d'antenne du terminal mobile est également requis car il est utilisé pour le localiser.

Pour permettre l'exécution des algorithmes de type Viterbi, le serveur a recours aux informations d'historique des clients, ainsi qu'à des données de topologie, contenues par la classe **Area**. Cette classe contient les données afférentes aux zones convexes de l'espace, et les points de liaison entre ces zones. Ces données permettent l'utilisation de l'algorithme de type Viterbi dans un ensemble infini de points, en particulier l'espace cartésien à trois dimensions.

Le serveur doit également conserver des données sur les points d'accès. En particulier, pour appliquer le FBCM, il est nécessaire de connaître les coordonnées des points d'accès, leurs puissances d'émissions ainsi que leurs gains d'antennes respectifs. Le rôle de la classe **AccessPoint** est de stocker ces informations et de permettre de les consulter, ainsi que de calculer et obtenir leurs indices de Friis respectifs.

La partie point de référence est gérée par les classes **ReferencePoint** et **Measurement**. La classe **referencePoint** contient les coordonnées du point de référence et un ensemble de mesures, contenues dans des objets de la classe **Measurement**. On y trouve les fonctions de calcul de la distance euclidienne dans l'espace du signal et des fonctions d'accès utiles à la recherche de mesures par rapport à un point d'accès. Comme une mesure est relative à un point d'accès mais contient plusieurs valeurs, la classe **Measurement** contient une fonction pour obtenir la moyenne des puissances observées et divers fonctions de manipulation des données.

Une dernière classe est utilisée par la classe **Server**. Il s'agit de la classe **Point** qui définit

un point dans l'espace cartésien. Cette classe possède des fonctions utiles, en particulier le calcul de la distance entre deux points de l'espace.

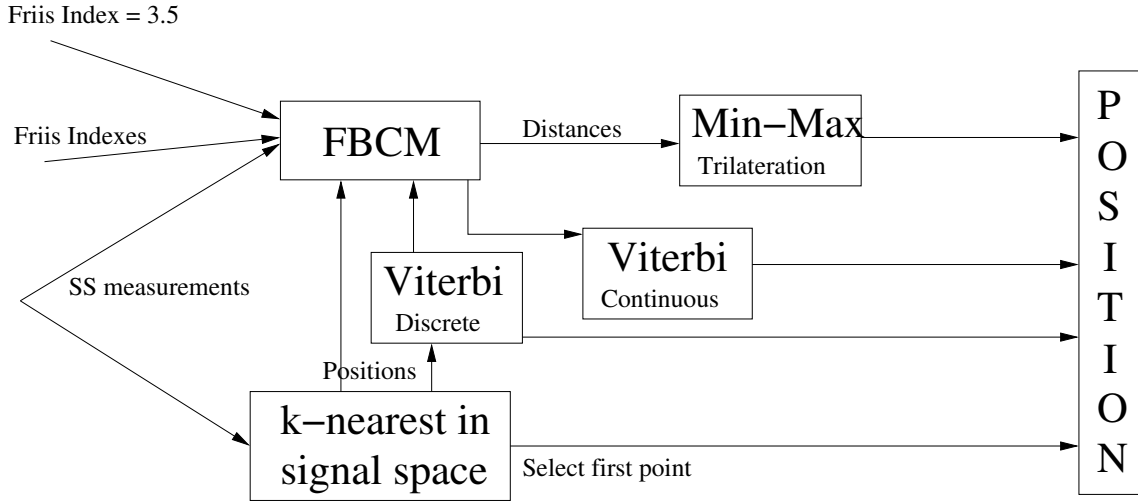


FIG. 5.3 – Architecture logicielle du service de gestion de la mobilité : algorithmes.

La figure 5.3 illustre la présence des algorithmes implantés dans le serveur de géolocalisation. L'algorithme **k-nearest**, qui sélectionne les k points de référence les plus proches d'une mesure dans l'espace du signal, est commun au FRBHM et à la méthode uniquement basée sur les points de référence. La fonction **FBCM** permet de calculer les distances entre les points d'accès reçus et le terminal mobile. Puis, elle calcule la position du terminal mobile par trilatération. Nous employons l'algorithme du min-max pour résoudre la trilatération. Remarquons que la fonction FBCM permet de déterminer une position suivant l'algorithme d'Interlink Networks : il suffit d'utiliser un indice de Friis égal à 3,5 pour tous les points d'accès. Combiner les approches FBCM et k plus proches voisins.

La mesure des puissances

La mesure peut être effectuée de deux façons. D'une part, le terminal mobile peut mesurer les puissances des signaux qu'il reçoit. D'autre part, les mesures peuvent être effectuées par l'infrastructure, en l'occurrence, les points d'accès. La mesure effectuée par le terminal mobile s'avère utile si l'on n'a pas accès au logiciel des points d'accès. De plus, elle sera ultérieurement requise pour obtenir un système de géolocalisation exécuté uniquement sur le mobile, et en mesure de se calibrer et se déployer indépendamment de l'infrastructure du réseau. La mesure effectuée par l'infrastructure permet de simplifier la mise en œuvre de la géolocalisation si l'on peut modifier le système logiciel des points d'accès. De plus, elle permet la mesure des puissances de tous les terminaux, par exemple dans une optique de détection d'intrusion.

La figure 5.4 expose le fonctionnement de la mesure lorsqu'elle est effectuée par le terminal mobile. Le terminal mobile s'appuie sur l'émission périodique des **beacons** (balises) par les points d'accès. Quand le terminal mobile souhaite obtenir un ensemble de mesures des puissances, il initie un décompte. Le décompte a une durée paramétrée en fonction de l'usage : court pour un

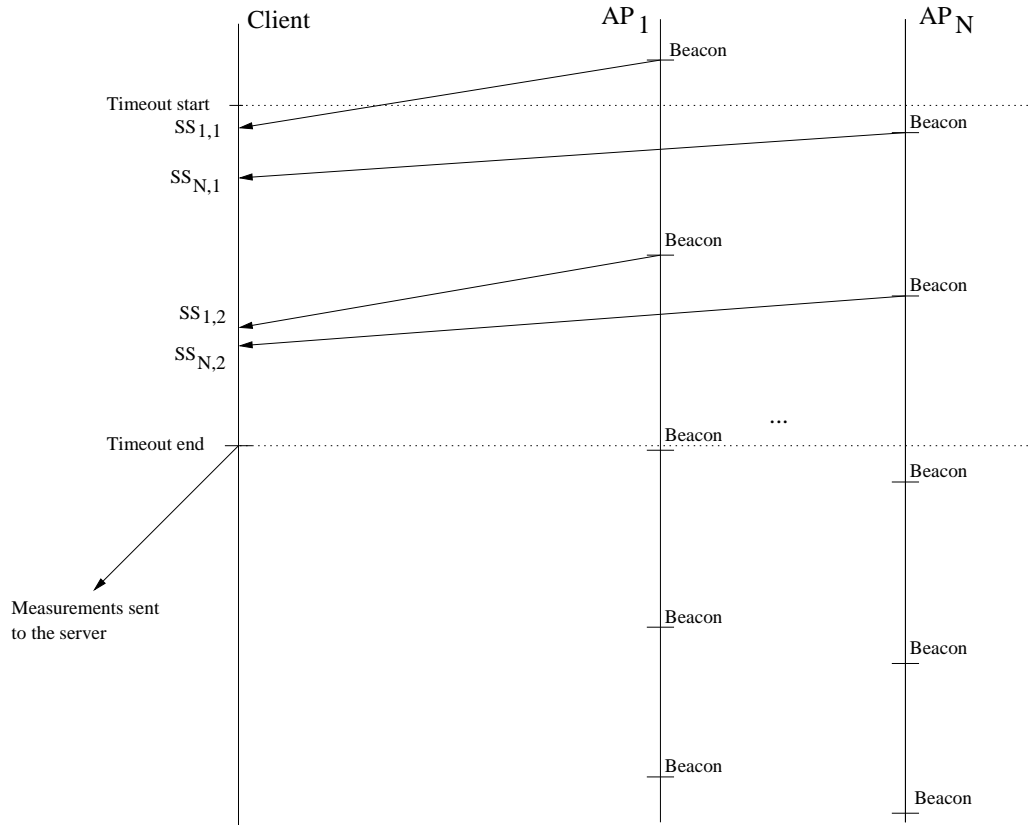


FIG. 5.4 – Mesure effectuée par le terminal mobile.

terminal mobile en mouvement, long pour un terminal mobile statique, par exemple lors de la calibration du FBCM et du FRBHM. Tant que le décompte n'est pas terminé, la puissance de toutes les balises reçues par le terminal mobile est mesurée. La valeur de la mesure, associée à l'adresse MAC du point d'accès, est stockée dans une liste. Quand le décompte est écoulé, la liste est envoyée au serveur de géolocalisation dans le paquet de requête de position du terminal mobile.

La figure 5.5 décrit la mesure effectuée par l'infrastructure. Trois acteurs sont requis : le terminal mobile, un ensemble de points d'accès et un serveur d'agrégation des mesures. La mesure est initiée lors de l'émission par le client d'un paquet de requête de position. Le paquet de positionnement contient un numéro de séquence unique, dans notre cas, la date au format unix. Ce type de paquet est reconnu par les points d'accès. Quand un paquet de ce type est reçu par un point d'accès, ce dernier mesure la puissance du signal transportant le paquet et l'envoie au serveur d'agrégation. Le rôle du serveur d'agrégation est de fusionner les mesures issues de plusieurs points d'accès. Pour ne pas attendre infiniment des paquets de tous les points d'accès, un décompte est utilisé. Le décompte s'initialise à la réception de la première mesure pour un couple terminal mobile-séquence. Pendant le décompte, toutes les mesures correspondant au même couple sont ajoutées à la liste. Quand le décompte s'arrête, le paquet de requête de position est formé et envoyé au serveur. Il contient l'identifiant du terminal mobile et l'ensemble des mesures.

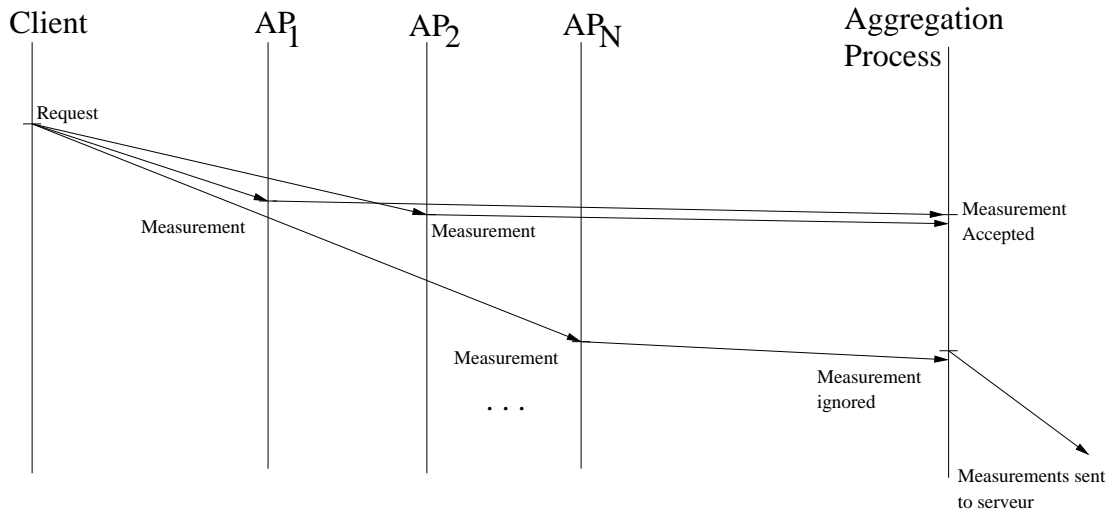


FIG. 5.5 – Mesure effectuée par l'infrastructure.

Utilisation

Ces architectures logicielle et matérielle nous ont permis de réaliser les expérimentations décrites dans la section 2.5. En particulier, les 2 derniers scénarios décrits ont été testés sur ce système après l'avoir déployé à Numérica. Nous y avons mesuré des points indépendants ainsi qu'une trajectoire de terminal mobile. Nous avons évalué la précision de plusieurs systèmes de géolocalisation et obtenu les résultats exposés dans la sous-section 2.5.3.

Conclusion

Dans ce chapitre, nous avons présenté des mises en situation de la gestion de la mobilité au sein d'applications. Les deux usages ciblés par la gestion de la mobilité sont liés à ses deux grandes problématiques : la géolocalisation et la prédiction de la mobilité. La géolocalisation permet de fournir des services relatifs au contexte du terminal mobile. Elle permet également de détecter des intrusions et de les localiser lorsque la partie sans-fil d'un réseau est compromise. La prédiction de la mobilité permet d'anticiper les déplacements des terminaux mobiles. Elle est utile dans des applications qui doivent assurer la continuité des services en situation mobile. Elle est également exploitable pour anticiper les congestions dues à une forte concentration de terminaux mobiles dans un réseau sans-fil.

Les travaux d'implantation de nos contributions sont actuellement en cours de réalisation et permettent de tester les algorithmes proposés. Il ne s'agit pour l'instant pas d'une application complète. Cependant, améliorer la gestion des erreurs et l'interfaçage entre les processus et avec d'autres applications permettra d'en faire un intergiciel exploitable de gestion de la mobilité.

Conclusion

Je soutiens la thèse que la gestion de la mobilité dans un réseau Wi-Fi s'appuie sur la géolocalisation et la gestion pro-active des procédures de *handover*, basée sur la prédiction de la mobilité. La géolocalisation est effectuée par un système hybride, combinant les approches discrètes et continues. La gestion pro-active du *handover* est rendue possible par l'apprentissage des pérégrinations des terminaux mobiles et leur représentation dans un modèle statistique. Le modèle est exploité pour prédire où un terminal mobile va se diriger prochainement et ainsi préparer sa procédure de handoff.

Dans la première partie, nous avons étudié les travaux relatifs à la géolocalisation de terminaux mobiles dans un réseau Wi-Fi. Nous en avons déterminé les modes de fonctionnement et une classification relatives aux modèles sous-jacents. Les deux modèles principaux sont l'utilisation d'une cartographie des puissances et le calcul de la position basée sur la propagation des ondes radio. Nous avons identifié les contextes pertinents en fonction des modèles de la classification. Alors que la cartographie des puissances se révèle fastidieuse à installer, du fait du nombre de mesures requises, les modèles de propagation se révèlent moins précis mais plus rapide à installer.

En nous basant sur ces deux constatations, nous avons proposé plusieurs modèles de géolocalisation. Le modèle de Friis calibré (FBCM) permet de calibrer la relation de Friis pour l'adapter à l'intérieur d'un bâtiment. Les modèles hybrides (FRBHM) obtiennent une précision de géolocalisation d'environ 3 mètres, tout en se satisfaisant d'un faible effort de calibration et de paramétrage. Ce modèle est applicable à l'intérieur des bâtiments car il tient compte de la topologie et des possibilités de déplacement des terminaux mobiles. Par ailleurs, nous avons mis en évidence qu'un modèle uniforme de propagation du signal est inapplicable dans un bâtiment. Ce fait est dû à la topologie hétérogène rencontrée dans les bâtiments.

Dans la seconde partie, nous avons traité la prédiction de la mobilité. Nous avons envisagé la prédiction de la mobilité dans sa définition d'une politique de *handover*. Nous avons étudié les modèles d'apprentissages tels que les chaînes de Markov, les modèles de Markov à états cachés et les réseaux bayésiens. Ces modèles ont tous été largement utilisés dans des problématiques de prédiction de phénomènes décomposables en états et transitions. Nous nous sommes intéressés particulièrement aux modèles de Markov et aux réseaux bayésiens. Les modèles de Markov ont été largement exploités dans la prédiction des visites de pages web. Les réseaux bayésiens ont été appliqués à la prédiction de la mobilité et sont principalement utilisés dans l'étude de divers systèmes pour lesquels plusieurs paramètres dépendants sont observables.

Dans un premier temps, nous avons construit un modèle composé de plusieurs degrés de modèles de Markov. Nous avons inclus à ce modèle un système de seuil de déclenchement de la

politique de prédiction de la mobilité. Ce modèle permet d'atteindre une proportion supérieure à 70% de prédictions correctes pour un seuil égal à 50%. Ce seuil garantit une seule position vers laquelle une politique de *handover* sera appliquée. Dans un second temps, comme la mobilité et la diffusion de contenus multimédia doivent tenir compte de la dimension temporelle, nous avons ajouté la variable *temps* au modèle. Le temps est défini par le délai nécessaire avant de déclencher la préparation du *handover*. La précision est peu altérée par le délai si on le choisit bien.

De ces deux parties, nous avons premièrement pu conclure que la géolocalisation de terminaux Wi-Fi en intérieur, bien qu'elle soit ardue, est possible. Beaucoup de paramètres entrent en jeu, tels que la topologie qui altère plus ou moins la précision des modèles de propagation. La conclusion la plus intéressante est la possibilité de concevoir un système de géolocalisation qui ne nécessite aucune information sur son environnement et qui se calibrera de lui même au cours du temps. Deuxièmement, à la lumière des résultats en prédiction de la mobilité, nous avons déterminé que prédire dans un futur proche les mouvements des terminaux mobiles est possible. L'exploitation des informations de prédiction reste ensuite à la discrétion des applications qui requièrent cette information.

Perspectives

L'utilisation des modèles de Markov à états cachés pourrait également permettre la fusion de la cartographie des puissances et de la prédiction de la mobilité au sein d'un unique modèle. En effet, dans un modèle de Markov à états cachés, il existe un ensemble d'observations, un ensemble d'états et des probabilités pour chaque observation de correspondre à un état donné. Le reste du modèle (matrice de transition et états initiaux) reste identique à un modèle de Markov standard. Dans notre contexte, les observations sont les mesures de puissance des signaux, les états sont les positions des points de référence. Dans cette optique, les probabilités de correspondance entre les observations et les états sont déterminables par un procédé semblable au calcul employé par le système HORUS [22].

Concernant les observations des modèles de géolocalisation, les données utilisées et les combinaisons de paramètres (nombre de points d'accès et densité des points de référence) sont riches. Cependant, il serait souhaitable de réitérer les mêmes manipulations dans un ou plusieurs autres bâtiments afin de confirmer ou modifier les conclusions tirées. En effet, le lieu de nos tests, Numérica, a une topologie très hétérogène qui ne se plie pas particulièrement à l'usage de modèles basés sur la propagation des ondes radio. Généraliser les observations dans des contextes topologiques différents, c'est-à-dire des bâtiments différents, permettrait de compléter les observations décrites dans ce document.

Dans la problématique de la localisation en intérieur, nous avons identifié 3 sujets qui méritent d'être approfondis. Il s'agit de proposer un système de positionnement sans aucune initialisation pour que celui-ci puisse se déployer automatiquement, tout en intégrant une précision au moins équivalente aux systèmes de positionnement cartographiques. Le deuxième sujet concerne l'extension des systèmes de positionnement intérieur à coordonnées relatives en système à coordonnées absolues en les associant aux GNSS. Le dernier sujet vise à concevoir un système de positionnement combiné permettant d'utiliser indifféremment des points de repère issus de deux technologies différentes, GNSS et Wi-Fi, et participant à un unique calcul de position.

Par ailleurs, concernant la prédiction de la mobilité, nous avons cherché à prédire la position d'un terminal mobile à son prochain mouvement. La question que nous nous posons est de savoir s'il est également possible, avec une précision suffisante, de déterminer des déplacements au delà du prochain. Cette fonction est théoriquement possible, en multipliant la matrice de transition du modèle par elle-même autant de fois que nous souhaitons proposer une probabilité future. Cependant, il est nécessaire d'observer la dégradation des probabilités calculées ainsi et son impact sur la précision de la prédiction.

Quatrième partie

Annexes

Annexe A

Notions fondamentales

Introduction

Dans ce chapitre, les notions fondamentales pour la compréhension de nos travaux sont présentées. Nous traitons premièrement la propagation des ondes radio, un médium de transmission dominant dans les liaisons sans-fil. Nous y présentons les unités de mesure utilisées et les phénomènes physiques qui s'appliquent aux ondes lors de leurs déplacements. Nous décrivons ensuite les techniques usuelles de calcul de la position, c'est-à-dire la trilatération et la triangulation. Puis nous présentons la norme IEEE 802.11, à laquelle nous nous intéressons particulièrement dans le cadre de nos travaux. Pour finir, nous présentons des modèles de représentation des données. Ces modèles sont les chaînes de Markov, les modèles de Markov à états cachés et les réseaux bayésiens.

A.1 Ondes radio

Dans cette section, nous présentons différentes causes physiques de l'affaiblissement de la puissance du signal radio lors de sa propagation. Par la suite, nous nommerons l'affaiblissement de la puissance du signal radio simplement *affaiblissement du signal*. La compréhension des causes d'affaiblissement du signal est nécessaire pour bâtir un modèle qui exprimera la distance entre un récepteur et un émetteur à partir de l'affaiblissement du signal entre ces deux appareils.

Premièrement, nous définissons en guise de rappel les unités de mesure employées dans l'étude de la puissance du signal, c'est-à-dire le décibel-milliwatt (dBm) et le décibel isotrope (dBi). Ce dernier caractérise les antennes. Deuxièmement, nous exposons les causes d'affaiblissement du signal liées à la trajectoire des ondes radio. Ces causes sont la diffraction, la réflexion et la réfraction. Troisièmement, nous décrivons les causes d'affaiblissement du signal liées à la distance parcourue par l'onde radio à partir de son émetteur.

A.1.1 Unités de mesure du signal

La puissance d'un signal se mesure généralement en Watt. Quand les quantités mesurées sont très petites ou très grandes, on utilise le décibel ramené au milliwatt (dBm). Si P_{mW} est la

puissance du signal exprimée en milliwatts, alors la puissance du signal exprimée en dBm est

$$P_{dBm} = 10 \log(p_{mW})$$

Le dBm est l'unité de puissance en échelle logarithmique. Il permet de simplifier les calculs du fait des propriétés du logarithme et de faciliter la représentation des valeurs exprimées en dBm.

Concernant les antennes, le pendant du dBm est le dBi (décibel isotrope). Il exprime le gain d'antenne, c'est-à-dire la concentration du signal par l'antenne, par rapport à une antenne isotrope. L'antenne isotrope est une notion théorique d'une antenne qui prendrait la forme d'un point et dont le rayonnement serait homogène autour d'elle. Dans la pratique, une antenne concentre toujours ses émissions dans un volume plus restreint qu'une sphère parfaite. Par conséquent, l'énergie des portions de l'espace hors de la couverture de l'antenne est redirigée dans la couverture de l'antenne. Cela a pour effet d'émettre dans la zone de rayonnement une puissance isotrope rayonnée équivalente (la) supérieure à la puissance émise par une antenne isotrope. Plus l'antenne est directionnelle, plus la concentration de l'énergie est grande et plus le gain est important.

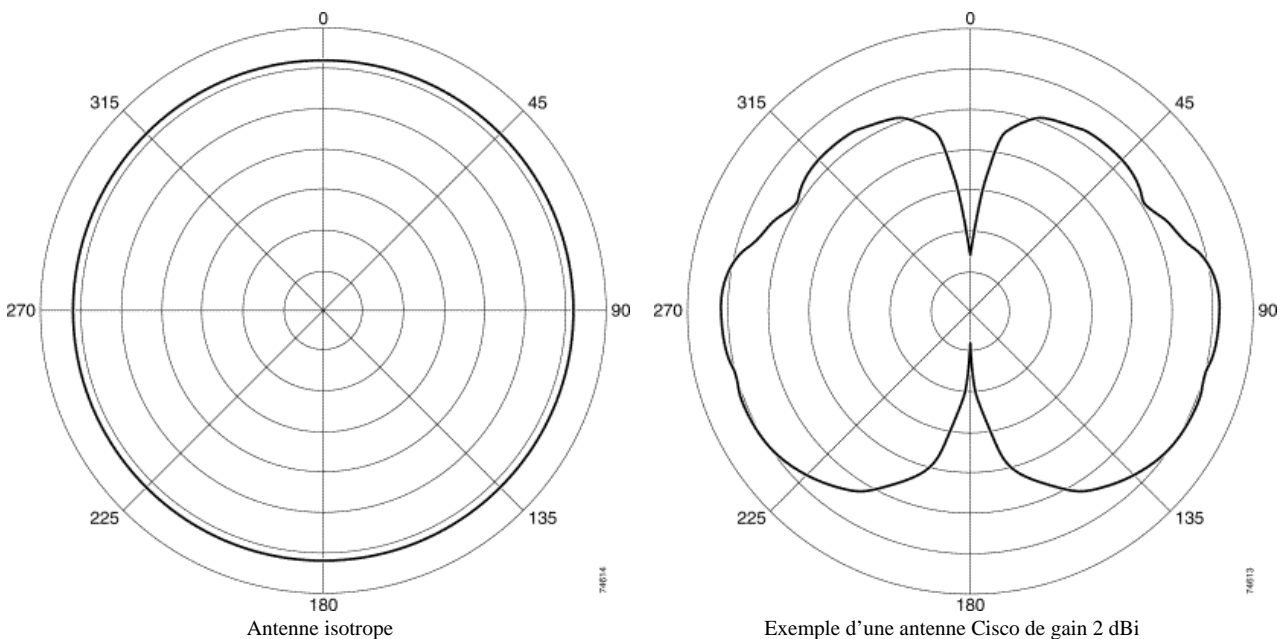


FIG. A.1 – Diagrammes de couverture d'antennes.

Le gain de l'antenne se définit donc comme la capacité de l'antenne à concentrer le signal (cf. fig. A.1). Il dépend également de la fréquence : plus la fréquence est élevée, plus le gain augmente pour une antenne donnée. La valeur du gain peut s'appliquer à n'importe quelle puissance, qu'elle soit exprimée en dB ou en dBm.

Il est intéressant d'observer que le gain est identique en émission et en réception. Il est égal à la somme du gain de l'émetteur et du gain du récepteur. En effet, le gain d'antenne du récepteur lui permet de concentrer les signaux environnant pour mieux les capter alors que

le gain d'antenne de l'émetteur lui permet de rendre le signal plus puissant dans sa zone de rayonnement.

A.1.2 Phénomènes de trajectoire

Lorsqu'une onde traverse un milieu et rencontre un obstacle, elle est déviée partiellement ou totalement. Ce changement de trajectoire a pour effet de modifier la puissance du signal derrière les obstacles et à créer des chemins multiples lors de la réflexion des ondes sur des surfaces.

Réflexion

Les ondes hertziennes sont soumises au phénomène de réflexion [52]. Il est semblable à la réflexion de la lumière mais a lieu sur la plupart des matériaux. Ce phénomène a pour effet de créer des chemins multiples d'une onde à partir d'une seule et même source. Ces chemins multiples ont pour effet la réception multiple d'un même signal. La figure A.2 illustre le rayonnement d'un signal à partir de l'émetteur et la réflexion des ondes sur des surfaces. Le récepteur reçoit plusieurs fois, avec un déphasage, les différentes ondes issues de l'émission d'origine. Il reçoit des ondes qu'il n'aurait pas reçues sans qu'elles soient réfléchies. Le déphasage est la conséquence des longueurs variables des différents chemins qui induisent des délais entre les réceptions multiples. Le problème des réceptions multiples est résolu en ignorant les trames reçues dans un temps supérieur à un délai fixé en fonction du débit des données. Ces aspects sont traités par l'appareil Wi-Fi et son firmware.

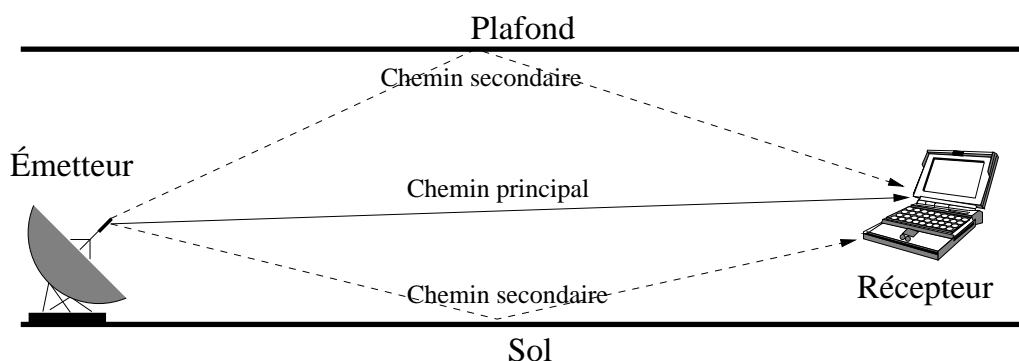


FIG. A.2 – Réflexion d'une onde sur les parois d'un bâtiment.

Les réflexions ont lieu sur différents éléments de la topologie. Les plus courants sont le sol, le plafond et les murs dans un bâtiment.

Réfraction

Comme la lumière dans les milieux transparents, les ondes radio sont sujettes à réfraction quand elles pénètrent dans un milieu différent. Ceci est particulièrement vrai dans la propagation ionosphérique¹³. La réfraction modifie la trajectoire d'une onde. Ce changement est déterminé par un indice de réfraction. Le dioptré est la surface de séparation entre les deux milieux. Nous

¹³La ionosphère est la dernière couche atmosphérique.

ne tenons pas compte dans notre problématique de la réfraction car elle est difficile à quantifier et entraîne des calculs longs.

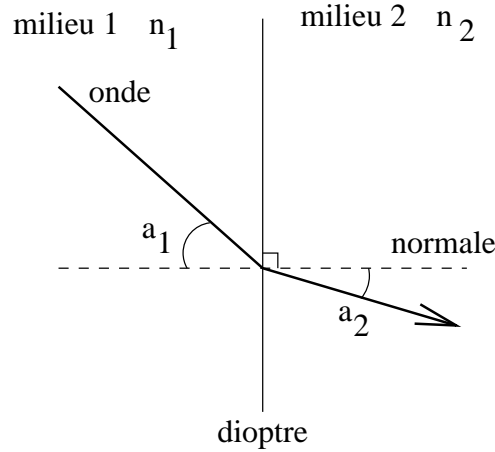


FIG. A.3 – Réfraction d’une onde pénétrant dans un milieu différent.

La figure A.3 illustre le phénomène de réfraction. La réfraction est déterminée par les règles de Snell-Descartes. Le dioptr est la surface entre les deux milieux au sein desquels l’onde circule. La loi de Snell-Descartes exprime la réfraction par la formule suivante :

$$n_1 \sin(a_1) = n_2 \sin(a_2)$$

où a_1 et a_2 sont les angles d’incidence de l’onde dans les milieux 1 et 2 et n_1 et n_2 sont les indices de réfractions des milieux 1 et 2.

Diffraction

C’est ce phénomène qui est pris en compte dans la définition de la zone de Fresnel, mais celle-ci ne nous est pas utile au sein d’un bâtiment. Cette notion est par contre très utile dans l’établissement de liens radio extérieurs où il permet de définir une zone devant rester vierge d’obstacles entre deux antennes.

La diffraction est liée à la nature des ondes. Si on considère que l’onde se déplace en ligne droite (cf. fig. A.4(a)), on fait une approximation qui ignore la diffraction. Dans la réalité, l’onde se propage de la façon décrite dans la figure A.4(b).

Ce phénomène permet d’obtenir un signal derrière des obstacles infranchissables pour une onde radio. L’onde qui se diffuse derrière l’obstacle est cependant affaiblie. En effet, après passage par l’obstacle, il ne reste que l’onde qui n’a pas été stoppée par l’obstacle. On explique la diffraction par la composition de l’onde en une multitude de petites ondelettes. Sur le front d’onde, chaque point agit comme la source d’un deuxième front d’onde. C’est le principe de Huygens [52] qui a permis l’explication et la compréhension de la diffraction.

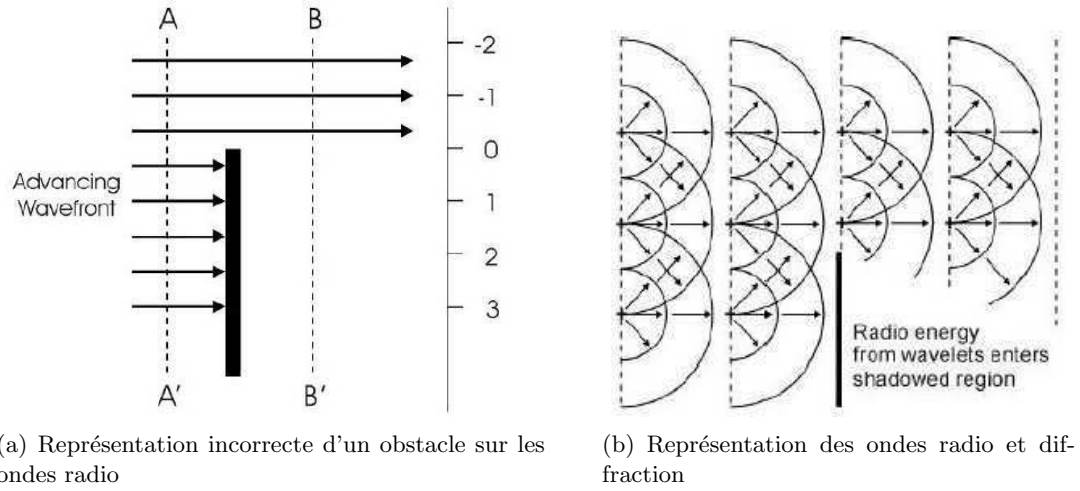


FIG. A.4 – Phénomène de diffraction.

A.1.3 Atténuation : la formule de Friis

Une onde radio perd sa puissance lorsqu'elle se déplace dans un milieu. Une partie de la puissance émise se dissipe sous forme d'énergie thermique au fur et à mesure du déplacement. L'atténuation de l'onde porteuse du signal est fonction de la distance parcourue par cette onde, et du milieu traversé. Même dans l'atmosphère dénuée d'obstacles, l'onde s'affaiblit. Elle s'affaiblit aussi, de manière plus importante, en traversant d'autres milieux.

La perte de puissance est en relation avec la distance parcourue par cette onde, qui s'affaiblit au fur et à mesure qu'elle se déplace. Cette perte dans un environnement sans obstacle, c'est-à-dire dans notre atmosphère, a déjà été mise en équation, aboutissant à la relation de Friis, qui est la suivante :

$$\frac{P_R}{P_T} = G_R G_T \left(\frac{\lambda}{4\pi d} \right)^2$$

avec :

- P_R et P_T correspondant respectivement à la puissance reçue et à la puissance d'émission (en Watt) ;
- G_R et G_T correspondant respectivement au gain d'antenne du récepteur et au gain d'antenne de l'émetteur ;
- λ correspondant à la longueur d'onde du signal (en mètre) ;
- d correspondant à la distance entre l'émetteur et le récepteur (en mètre).

La figure A.5 est la courbe de l'atténuation du signal Wi-Fi en dBm en fonction de la distance en mètres entre l'émetteur et le récepteur. Comme l'exprime la relation de Friis, la puissance reçue diminue quand la distance entre l'émetteur et le récepteur augmente. La diminution de la puissance dépend aussi de la fréquence $f = \frac{v_{onde}}{\lambda}$ du signal. Dans le cas du signal radio porteur des informations de la norme 802.11, la longueur d'onde est de l'ordre de 12 cm. La vitesse de l'onde dans l'atmosphère est d'environ $v_{onde} = 3.10^8 m.s^{-1}$. Si la relation de Friis se limite à modéliser la propagation d'une onde radio dans l'atmosphère, la perte de puissance s'applique également lors du franchissement d'obstacles. Cet affaiblissement est supérieur à celui lié à l'atmosphère.

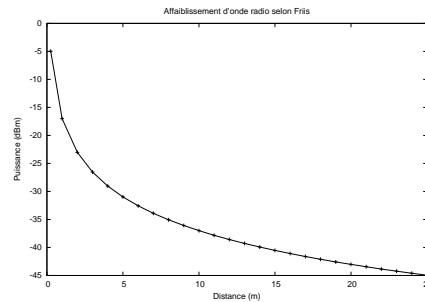


FIG. A.5 – Atténuation du signal suivant la formule de Friis.

Comme les obstacles sont le plus souvent des murs, on parle alors d'absorption du signal par les obstacles (contrairement à l'atténuation du signal en fonction de la distance). Cependant, ces deux phénomènes ont la même origine.

A.2 Calcul de positionnement

Déterminer la position d'un objet nécessite des méthodes de calcul de position par rapport à des objets dont les coordonnées sont connues. Dans cette section, la triangulation et la trilatération sont décrites et analysées.

A.2.1 Triangulation

La triangulation est une méthode attribuée au mathématicien Thalès. Thalès développa la méthode appelée triangulation pour calculer la position d'un navire. La figure A.6 illustre la construction géométrique de la triangulation. Deux observateurs A et C sur le rivage observent chacun l'angle de vue par rapport au navire B . Ils mesurent les angles a et c créés respectivement par les directions (AB) et (CB) avec le segment de droite $[AC]$. Comme la distance AC est mesurable et par conséquent connue, le triangle ABC peut être construit. Connaissant AC et les angles a et c adjacents, il est possible de déduire AB et BC , qui sont les distances du navire aux points A et C .

La triangulation est utilisée pour calculer la position des planètes et des étoiles en astronomie. La mesure des positions des astres par la triangulation entraîne des problèmes géométriques qui découlent du rapport de la distance entre deux points sur la Terre comparée à la distance avec un astre lointain. Il en découle que les angles a et c tendent les deux vers 90° . Dans ce cas, les erreurs de mesure des angles ont un impact énorme sur la précision du positionnement. Par exemple, si les angles a et c sont égaux à 60° chacun, l'erreur de calcul de la distance est d'environ AB est de l'ordre de 1,5 %. Si les angles a et c sont égaux à 89° chacun, l'erreur obtenue est alors d'environ 100 % de la distance AB . Pour résoudre ce problème, il est possible d'utiliser deux mesures d'angles à des temps différents car nous connaissons la trajectoire de la Terre dans la galaxie. L'usage du mot triangulation dans le cas du Global Positioning System (GPS, cf. paragraphe 1.4.1), ou des techniques semblables exposées dans le chapitre 1, est un abus de langage et ne présente pas la technique originelle.

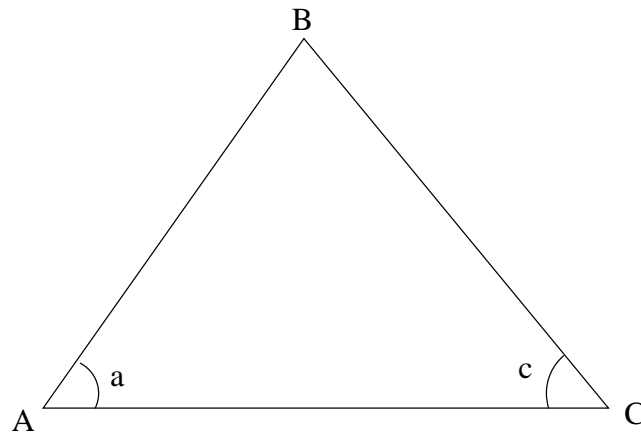


FIG. A.6 – Principe de la triangulation

A.2.2 Trilatération

La trilatération consiste à déterminer la position d'un point de coordonnées inconnues grâce aux distances entre ce point et des points dont les coordonnées sont connues à tout instant. Connaissant un point, ses coordonnées et sa distance par rapport au point dont on recherche les coordonnées, on peut déterminer un cercle (respectivement une sphère) dans le plan (respectivement l'espace) centré(e) sur le point lui-même et de rayon la distance par rapport au point de coordonnées inconnues (voir fig. A.7). Dans le cas présent, 3 cercles de centres respectifs C_1 , C_2 et C_3 et de rayons variables ont pour intersection le point M . Le rayon de chaque cercle est la distance, de son centre au mobile, induite par la puissance mesurée. Le point M est la position du terminal mobile calculée par trilatération. Les cercles (respectivement les sphères) centré(e)s sur les points de coordonnées connues et de rayon égal à la distance séparant le point et l'objet dont la position est à définir sont appelés *cercles de distance* (respectivement *sphères de distance*).

On en déduit un système d'équations à N inconnues pour un espace à N dimensions, généralement 2 ou 3. La résolution de ce système nous permet alors d'obtenir les coordonnées du point qui nous intéresse. Dans le cas où le nombre de cercles, donc de points de référence, est supérieur à la dimension de l'espace de géolocalisation, le système est résolu par un ensemble de $N + a$ équations dont un nombre a d'équations seront linéairement dépendantes.

Il s'agit d'une situation idéale. Dans la réalité, les distances induites par les puissances mesurées ne peuvent pas être considérées comme exactes. On obtient donc des cercles qui possèdent deux à deux, soit zéro, soit un ou deux points d'intersection. Il est alors possible d'approximer la position du terminal mobile par une région de l'espace de géolocalisation suffisamment restreinte pour l'application envisagée, un cube de 5 mètres de côté pour la localisation d'un véhicule par exemple. Dans ce cas, la résolution pour un nombre de points de référence supérieur à la dimension N de l'espace de géolocalisation doit permettre de générer une solution optimale vis-à-vis de tous les cercles. Il est possible d'utiliser les données au-delà des N premières pour affiner un premier résultat ou de calculer la position en s'approchant au plus près de l'ensemble des cercles de distance.

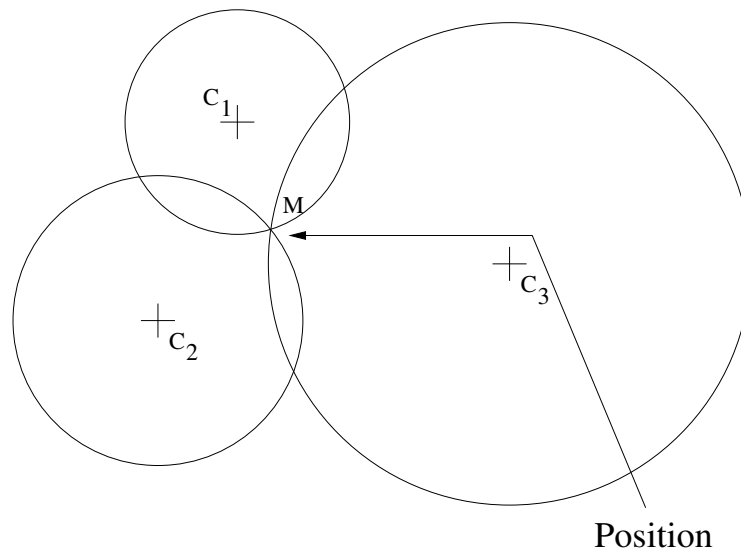


FIG. A.7 – Principe de la trilatération.

A.3 La norme 802.11

La norme IEEE 802.11 a donné lieu à deux générations de réseaux sans fil [53], les réseaux Wi-Fi 2 qui travaillent à la vitesse de 11 Mbit/s, et les réseaux Wi-Fi 5, qui travaillent à 54 Mbit/s. Les premiers réseaux se fondent sur la norme IEEE 802.11b et les seconds sur la norme IEEE 802.11a.

Pour le réseau Wi-Fi 2, les fréquences choisies se situent dans la gamme des 2,4 GHz. Dans cette solution de réseau local par voie hertzienne, les communications peuvent se faire soit directement de station à station, mais sans qu'une station puisse relayer les paquets vers une autre station terminale, soit en passant par une borne de concentration que l'on appelle point d'accès ou AP (Access Point).

Dans cette section, nous présentons dans un premier temps les méthodes d'accès au médium de transmission dans la norme 802.11. Dans un second temps, nous présentons les révisions majeures de la norme 802.11 et leurs spécificités.

A.3.1 Accès au médium de transmission

Les méthodes pour accéder au support physique, décrites dans [54], sont au nombre de deux : le DCF (*Distributed Coordination Function*) et le PCF (*Point Coordination Function*). Nous présentons d'abord la méthode DCF et son extension par le mécanisme RTS/CTS utile dans certaines configurations topologiques. Ensuite la méthode PCF, moins fréquente sur les périphériques 802.11, est décrite.

DCF

La technique d'accès au support physique, qui est effectuée par un protocole, dit protocole MAC (*Medium Access Control*) puisque interne au niveau MAC, est la même pour tous les

réseaux Wi-Fi. De nombreuses options rendent toutefois sa mise en œuvre assez complexe. Ce protocole MAC se fonde sur le CSMA/CD, utilisé dans les réseaux Ethernet filaires.

La différence entre le protocole hertzien et le protocole filaire provient de la façon de détecter les collisions. Dans le cas filaire, on détecte les collisions en écoutant la porteuse, mais lorsque deux stations veulent émettre pendant qu'une troisième transmet sa trame, cela mène automatiquement à une collision. Dans le cas hertzien, l'écoute de la porteuse est impossible lors d'une émission car tout signal émis est beaucoup plus puissant que les signaux reçus des autres stations. La première caractéristique de la couche MAC dans la norme 802.11 est donc d'utiliser des acquittements pour détecter les collisions et permettre la retransmission des paquets perdus. Cependant, dans le cas des *broadcasts*, l'acquiescement devient impossible car il entraînerait des collisions. L'idée est d'attendre que le support soit libre pour émettre. Dans le cas où il y a collision, le paquet est retransmis après un temps aléatoire. Dans le cas de la norme 802.11, il est impossible de détecter la collision qui aurait lieu *pendant* l'émission. Il faut donc recourir à un autre système de résolution des collisions : le CSMA/CA pour *Collision Avoidance*. Dans le cas du CSMA/CA, lorsque la station a écouté la porteuse et que plus rien n'y circule, au lieu d'émettre de suite, chaque station qui veut émettre tire un *backoff* aléatoire. La différence entre deux valeurs données du *backoff* de deux terminaux étant supérieure à la durée de propagation du signal d'une station à une autre, une station qui a un temps d'attente plus long trouve le support physique déjà occupé et peut le détecter car elle n'émet pas encore, c'est ce qui fait que ce mécanisme évite les collisions.

Le mécanisme RTS/CTS

Si le CSMA/CA permet d'éviter les collisions, il existe cependant des cas où les collisions auront tout de même lieu, notamment dans le cas des nœuds cachés. Sur la figure A.8, on voit que les terminaux de part et d'autre du mur ne se détectent pas, pourtant, les deux peuvent communiquer avec le terminal central.

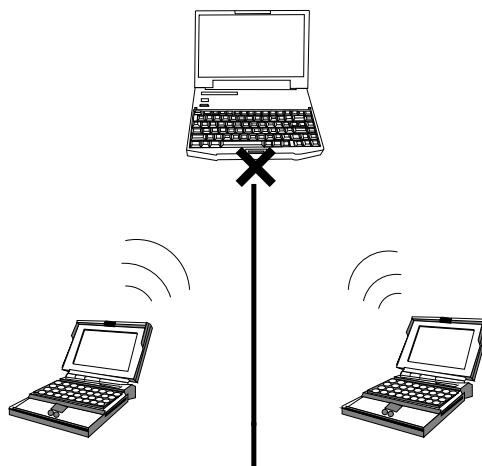


FIG. A.8 – Configuration des nœuds cachés

Dans ce cas, les deux émetteurs ne se voient pas et ne détectent pas leur activité respective.

Chacun croit donc que le canal est libre et émet dès qu'il a des données disponibles. Cependant, au niveau du récepteur, les chances de collisions sont très élevées. Il existe un mécanisme qui utilise des paquets de contrôle appelés *Request To Send* (RTS) et *Clear To Send* (CTS) et qui a été introduit dans [55]. Pour émettre, un mobile n'enverra pas immédiatement un gros paquet de données mais un paquet RTS de petite taille qui a donc moins de chances de provoquer une collision. À la réception du RTS, le destinataire répond par un paquet CTS envoyé à tout le voisinage. Les paquets RTS et CTS contiennent les informations nécessaires à la réservation du médium pour la transmission des données. Le mobile ayant envoyé le RTS sait donc, à la réception du paquet CTS, qu'il peut envoyer ses données.

Cette réservation de canal est implémentée grâce au *Network Allocation Vector* (NAV). Ce NAV indique dans chaque nœud le temps de réservation du canal de communication. Cette indication est indépendante de l'écoute du canal. On parle aussi de détection "logique" de porteuse. En effet, à la réception du paquet RTS, les mobiles non concernés savent combien de temps le canal va être occupé par la transmission.

PCF

Si le CSMA/CA permet d'éviter les collisions dans les réseaux 802.11 en utilisant un *backoff*, c'est également des unités de temps perdues correspondant à un débit effectif perdu. C'est pour cela qu'il existe aussi un système centralisé d'allocation de la bande passante. PCF est son nom. Son principe de base est de centraliser la gestion de l'accès au médium d'une cellule. C'est donc le point d'accès qui indique aux mobiles qui lui sont rattachés quand ils doivent émettre leurs paquets. Pendant tout le temps où le point d'accès impose l'ordre des transmissions, il n'y a pas de contention d'où le nom de *Contention Free Period*.

PCF conserve aussi la compatibilité avec les cartes qui n'implémentent pas PCF. Il existe donc dans chaque cycle PCF une période DCF pour ces dites cartes. En effet, le mécanisme PCF n'est pas implémenté sur la majorité des cartes 802.11.

A.3.2 Les révisions de la norme 802.11

L'élaboration de la norme 802.11 date de 1997 [56]. À cette date, 802.11 constitue une couche MAC commune à tous les WLAN : infrarouge et FHSS (Frequency Hopping Spread Spectrum) ou DSSS (Direct Sequence Spread Spectrum) à 2,4 GHz. Tous fonctionnent alors à 1 et 2 Mbps. L'accès OFDM (Orthogonal Frequency Division Multiplexing) est ajouté par la révision 802.11a de la norme en 1999. Cette révision définit une fréquence d'utilisation de 5 GHz et un débit maximum théorique de 54 Mbps. Elle est employée en particulier aux États-Unis. La même année, la norme 802.11b élargit l'utilisation de la fréquence de 2,4 GHz en DSSS à des débits de 5,5 et 11 Mbps. Puis, en 2003, 802.11g augmente le débit disponible à cette fréquence à un maximum théorique de 54 Mbps.

De nombreuses autres révisions de la norme 802.11 continuent d'être développées. En particulier, 802.11k vise à améliorer les mesures des ressources radio pour faciliter la gestion des WLAN 802.11. 802.11r concerne le *fast-roaming* entre AP. Il a été standardisé en Août 2008 avec un délai de 50 ms. Parallèlement, 802.11n, toujours au stade de draft 2.0, vise à augmenter les débits possibles à 540 Mbps théoriques en modifiant les couches physique et MAC, grâce à

l'utilisation de plusieurs antennes en émission et en réception (Multiple Input-Multiple Output, MIMO). Il est également intéressant de mentionner la norme en cours de définition 802.11v : elle intègre la gestion du *load-balancing* et également du *handoff*.

A.4 Modèles de Markov

Dans cette section, nous présentons les modèles de Markov et les modèles de Markov à états cachés. Alors que les premiers permettent de décrire un système en fonction de ses états et des probabilités de transition d'un état à un autre état, les modèles de Markov à états cachés permettent de prendre en compte l'inexactitude de l'observation de l'état d'un système.

Nous présentons dans un premier temps les chaînes de Markov à espace d'états discret et les variantes qui considèrent un historique supérieur à un unique état. Dans un second temps, nous décrivons le modèle de Markov à états cachés et les problématiques qui y sont liées concernant la correspondance entre les observations et les états sous-jacents du système.

A.4.1 Chaînes de Markov à espace d'états discret

Les chaînes de Markov sont des processus stochastiques sans mémoire. On peut définir un modèle de Markov par un ensemble d'états $E = \{e_1; e_2; \dots e_{\Omega(E)}\}$, une matrice de transition T dont chaque élément $t_{i,j}$ est la probabilité de transition de l'état e_i à l'état e_j . La somme des probabilités $t_{i,j}$ depuis un sommet e_i est égale à 1 :

$$\forall i \in [1; \Omega(E)], \sum_{j=0}^{\Omega(E)} t_{i,j} = 1$$

où $\Omega(E)$ est le cardinal de l'ensemble des états du modèle de Markov. La notion de processus sans mémoire tient au fait que, dans une chaîne de Markov, la probabilité de passage vers un état ne dépend que de l'état précédent. Si l'on cherche à modéliser des phénomènes dans lesquels la probabilité de passage vers un état est dépendante d'un historique plus long, il est possible d'utiliser des modèles de Markov d'ordres supérieurs à 1.

A.4.2 Chaînes de Markov d'ordres supérieurs à 1

Comme la chaîne de Markov est un processus sans mémoire, il est possible de considérer un historique plus long en le conservant comme label des états du modèle. Cela revient à ne plus représenter un état physique par un état du modèle. On considère à la place une succession d'états physiques comme un seul état du modèle [57].

A.4.3 Modèle de Markov à espace d'états caché

Les chaînes de Markov font l'hypothèse implicite que les états du système observé sont mesurables sans biais. Dans certains cas, cette assertion est fausse. Par exemple, considérons le découpage du plan en carrés de 50 mètres de côté. Chaque fois qu'un mobile transmet ses coordonnées, on cherche le carré dans lequel il est présent. Un mobile se déplaçant dans cet espace obtient sa position par l'intermédiaire du GPS. L'erreur de localisation du GPS est en moyenne

de 5 mètres. Il y a donc une probabilité non négligeable que le carré déterminé soit faux et que le mobile soit en réalité dans un carré adjacent. Le modèle de Markov à états cachés est la réponse à ce problème.

Définition

Le modèle de Markov à états cachés présenté par Lawrence R. Rabiner dans [58] existe depuis les années 60 et devient de plus en plus populaire grâce à la richesse de sa structure mathématique qui lui permet de poser les bases théoriques d'un grand nombre d'applications. Les chaînes de Markov sont composées d'états correspondant à des états physiques du système modélisé. La différence apportée par les modèles de Markov à états cachés est la possibilité de modéliser des systèmes dont tous les paramètres ne sont pas observables.

Plus formellement, un modèle de Markov à états cachés (*Hidden Markov Model*, HMM) est un double processus stochastique dans lequel existent des états non observables entre lesquels les transitions se font suivant une loi de probabilité, modélisable par un vecteur de transitions, pour chaque état, ce qui permet d'obtenir une matrice de transition. Cependant, les observations faites sont aussi des processus stochastiques différents pour chaque état du modèle.

Un modèle de Markov à états cachés est caractérisé par :

- N , le nombre d'états physiques, non visibles, du modèle ;
- M , le nombre d'observations possibles. Ces observations sont les phénomènes physiques observables en sortie du système à modéliser ;
- $A = \{a_{ij}\}$, la distribution de probabilité de transition d'un état i de N à un état j de N ;
- $B = \{b_j(k)\}$, la distribution de probabilité des observations où $\{b_j(k)\}$ est la probabilité que l'observation k corresponde à l'état j ;
- $\pi = \{\pi_i\}$, la distribution initiale de probabilité où π_i est la probabilité que l'état initial soit l'état i .

Une notation compacte d'un modèle de Markov à états cachés existe. On note alors le HMM $\lambda = (A, B, \pi)$. Le modèle de Markov à états cachés présente trois problématiques à résoudre :

- Connaissant une séquence d'observations $O = O_1, O_2 \dots O_T$ et un modèle $\lambda = (A, B, \pi)$, comment peut-on calculer $P(O|\lambda)$, la probabilité de cette séquence d'observations pour un modèle donné ?
- Connaissant une séquence d'observations $O = O_1, O_2 \dots O_T$ et un modèle $\lambda = (A, B, \pi)$, comment peut-on choisir une séquence d'états $Q = q_1, q_2 \dots q_T$ optimale dans un sens qui explique le mieux cette succession d'observations ?
- Comment peut-on ajuster les paramètres du modèle $\lambda = (A, B, \pi)$ pour maximiser $P(O|\lambda)$?

La problématique qui nous intéresse plus particulièrement est la deuxième car notre modèle doit nous permettre de reconstituer le parcours d'un mobile pour en prédire la suite. La résolution de ce problème n'est pas déterministe. Il existe en effet plusieurs solutions à ce problème et cette résolution tend à rechercher la séquence optimale d'états pour la séquence d'observations. La difficulté est de définir ce que l'on entend par optimal car les critères pour en juger sont nombreux. La première possibilité est de choisir pour chaque observation l'état qui est statistiquement le

plus probable. Ce type de résolution peut amener à construire une séquence invalide d'états car des transitions entre ceux-ci seraient de probabilité zéro.

La solution usuelle est d'utiliser l'algorithme de Viterbi pour calculer quelle séquence est la plus probable. L'algorithme de *Viterbi* cherche à maximiser la probabilité de la séquence d'états antérieurs pour l'état courant. Il est semblable à l'algorithme Viterbi-like décrit dans [15]. Dans le cas des modèles de Markov à états cachés, on calcule la probabilité des successions d'états correspondant à la succession d'observations. La solution retenue est la succession d'états qui maximise la probabilité. Cet algorithme est la solution à la seconde problématique soulevée par les modèles de Markov à états cachés. C'est également une solution à la prédiction de mobilité.

A.4.4 Algorithme de Viterbi

L'algorithme de Viterbi est utilisé dans la résolution de problèmes liés aux modèles de Markov à états cachés (HMM pour *Hidden Markov Model*). Il consiste à élire, à chaque itération de l'utilisation du HMM, un nombre paramétré d'états dont la probabilité de correspondre à l'observation courante est élevée. Ensuite, connaissant la probabilité que chaque état soit sous-jacent à l'observation, possédant un ensemble de sélections d'états et la probabilité de transition des états deux à deux, l'algorithme de Viterbi recherche la chaîne d'états la plus probable.

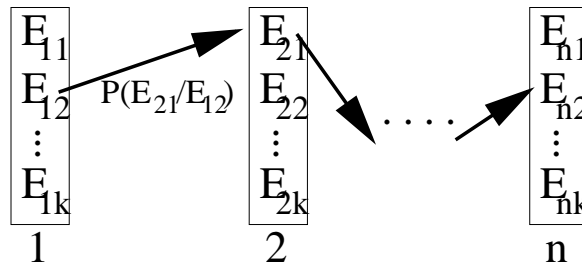


FIG. A.9 – Algorithme de Viterbi.

La figure A.9 représente les sélections 1 à N de K états choisis aux itérations 1 à N . Les flèches entre les sélections indiquent que le passage ne peut se faire que d'un état de la sélection i à un état de la sélection $i + 1$. Pour chaque chemin, on calcule le produit des probabilités de la correspondance de chaque état avec l'observation de l'itération correspondante et des probabilités de transition de chaque état vers son successeur, c'est-à-dire :

$$E_{final} = \left(\prod_{i=1}^N P(E_{i,j_i}/O_i) \right) \times \left(\prod_{i=1}^{N-1} P(E_{i+1,j_{i+1}}/E_{i,j_i}) \right)$$

où :

- ▶ E_{final} est la probabilité finale d'observer la séquence d'états $\{E_{0,j_0}, E_{1,j_1}, \dots, E_{N,j_N}\}$;
- ▶ P_{i,j_i} est le i -ième état de la séquence testée ;
- ▶ $P(E_{i,j_i}/O_i)$ est la probabilité que l'état E_{i,j_i} de la séquence corresponde à l'observation O_i de la séquence d'observations étudiées ;
- ▶ $P(E_{i+1,j_{i+1}}/E_{i,j_i})$ est la probabilité de transition depuis l'état E_{i,j_i} vers l'état $E_{i+1,j_{i+1}}$.

A.5 Inférence Bayésienne

Les réseaux bayésiens [59] sont des graphes qui représentent de façon compacte les relations entre les variables aléatoires d'un problème donné. Ces graphes permettent d'expliquer les causes d'un phénomène, de prédire un comportement en se basant sur des hypothèses connues et d'établir un diagnostic d'un phénomène. Les réseaux bayésiens sont utilisés en particulier dans les diagnostics médicaux et industriels, l'analyse de risques, la détection des spams et la fouille de données (*data mining*).

Le travail avec les réseaux bayésiens consiste dans un premier temps à définir le graphe de dépendance des variables aléatoires. La suite de cette section est illustrée par l'exemple suivant : les causes d'accident lors de l'utilisation d'une machine industrielle¹⁴. On sait que l'opérateur d'une machine risque de se blesser s'il l'utilise mal. Ses risques sont dépendants, entre autres paramètres, de son expérience du métier ainsi que de la complexité d'utilisation de la machine.

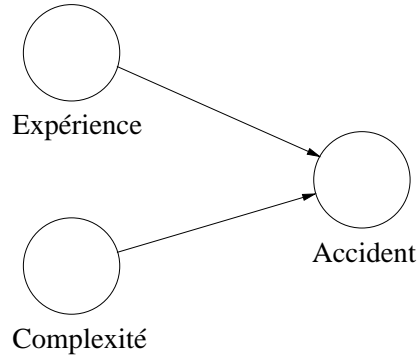


FIG. A.10 – Graphe de dépendance du réseau bayésien “risque d’accident”.

La figure A.10 présente le graphe de notre exemple. Dans le graphe, les nœuds “Expérience” et “Complexité” sont les parents du nœud “Accident”. Le nœud “Accident” est dépendant de ses nœuds parents. La probabilité d’observer un accident est dépendante de la complexité de la machine et de l’expérience de l’opérateur. Si on considère que l’opérateur est expérimenté ou ne l’est pas et que la machine est complexe ou ne l’est pas, les nœuds “Complexité” et “Expérimenté” prennent les valeurs “Vrai” (V) ou Faux (F) suivant une probabilité définie.

Complexité (C)	Expérience (E)	Accident (A)	Pas d’accident (\bar{A})
V	V	$P(A/C \cap E)$	$P(\bar{A}/C \cap E)$
V	F	$P(A/C \cap \bar{E})$	$P(\bar{A}/C \cap \bar{E})$
F	V	$P(A/\bar{C} \cap E)$	$P(\bar{A}/\bar{C} \cap E)$
F	F	$P(A/\bar{C} \cap \bar{E})$	$P(\bar{A}/\bar{C} \cap \bar{E})$

TAB. A.1 – Table des probabilités conditionnelles.

La probabilité d’observer un accident selon l’état du système est représentée dans le tableau A.1. Dans l’état initial, on détermine les probabilités des événements parents par des études sta-

¹⁴Exemple donné par Wikipedia, http://fr.wikipedia.org/wiki/Réseau_bayésien

tistiques ou des connaissances "experts". Les probabilités des événements enfants sont exprimées conditionnellement aux probabilités de leurs parents. On peut en déduire les probabilités a priori d'observer les valeurs prises par les enfants. En prenant connaissance de la valeur d'un parent à un moment donné, on modifie les conditions dans le système décrit. Alors les probabilités d'observer chaque valeur de l'enfant changent. En connaissant les valeurs pour suffisamment de parents, on obtient une idée précise de la probabilité d'observer chaque valeur de l'enfant.

Un élément favorable à l'utilisation n'est pas mis en valeur dans l'exemple. Il s'agit de la complexité du modèle établi par rapport à un modèle basé sur des probabilités jointes. Dans le cas de probabilités jointes, il est nécessaire de calculer les probabilités d'observer chacune des combinaisons des paramètres du système. Plus il y a de paramètres, plus le modèle contiendra d'informations. Dans le cas du réseau bayésien, seules les probabilités jointes des variables directement dépendantes entre elles sont établies. Ainsi, alors que l'utilisation de probabilités jointes mène à l'explosion du nombre de combinaisons étudiées, le réseau bayésien circonscrit l'ensemble des combinaisons au minimum nécessaire pour la résolution du problème.

Bibliographie

- [1] Michel Banâtre, Paul Couderc, Julien Pauty, and Mathieu Becus. Ubibus : Ubiquitous computing to help blind people in public transport. In *Proceedings of the 6th Int. Symposium on Mobile Human-Computer Interaction (MobileHCI 2004)*, pages 310 – 314, September 2004.
- [2] Nabil Jardak, Nel Samama, and Alexandre Vervish-Picois. Futurs GNSS et continuité intérieur-extérieur de la fonction de localisation. In *Actes des troisièmes journées Francophones Mobilité et Ubiquité (UbiMob'06)*, volume 1, pages 41–48, September 2006.
- [3] Joshua A. Tauber. Indoor Location Systems for Pervasive Computing. Massachusetts Institute of Technology. Area exam report., August 2002.
- [4] B.P. Crow, I. Widjaja, J.G. Kim, and P. Sakai. Ieee 802.11 wireless local area networks. *IEEE Communications Magazine*, 9(35) :116 – 126, September 1997.
- [5] Roy Want, Andy Hopper, Veronica Falc ao, and Johnathan Gibbons. The active badge location system. *ACM Transactions on Information Systems*, pages 91 – 102, 1992.
- [6] M. Fukumoto and M. Shinagawa. Carpetlan : A novel indoor wireless(-like) networking and positioning system. In *Proceedings of the 7th International Conference on Ubiquitous Computing (UbiComp 2005)*, pages 1–18, September 2005.
- [7] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location support system. pages 32 – 43. ACM press, August 2000.
- [8] Gaetano Borriello, Alan Liu, Tony Offer, Christopher Palistrant, and Richard Sharp. Walrus : Wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd int. conf. on Mobile systems, applications, and services (MobiSys 2005)*, pages 191 – 203. ACM Press, June 2005.
- [9] US Army Corps of Engineer, editor. *Engineering and Design - NAVSTAR Global Positioning System Surveying*. Number EM 1110-1-1003. Department of the Army, Washington, DC, July 2003.
- [10] Julien Caratori, Marc François, Nel Samama, and Alexandre Vervisch-Picois. Upgrade rns indoor positioning system in an office building. In *Proceedings of ION GNSS 2004*, pages 1959 – 1979, September 2004.
- [11] A. Ward, A. Jones, and A. Harper. A new location technique for the active office. *IEEE Personnal Communications*, 5(4) :42 – 47, October 1997.
- [12] S. Moix, C. Steiner, Q. Ladetto, and B. Merminod. Capteurs et analyse de signaux pour la navigation pédestre. *MPG*, pages pp. 512–516, August 2002.

- [13] Youngjune Gwon, Ravi Jain, and Toshiro Kawahara. Robust indoor location estimation of stationary and mobile users. In *Proceedings of the 23rd Conf. of the IEEE Communications Society (INFOCOM 2004)*, volume 2, pages 1032 – 1043, March 2004.
- [14] R. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen. A Probabilistic Approach to WLAN User Location Estimation. *International Journal of Wireless Information Networks*, 9(3) :155–164, July 2002.
- [15] Paramvir Bahl and Venkata N. Padmanabhan. RADAR : An in-building RF-based user location and tracking system. In *INFOCOM (2)*, pages 775–784, 2000.
- [16] P. Bahl, A. Balachandran, and V. Padmanabhan. Enhancements to the radar user location and tracking system. Technical report, 2000.
- [17] Asim Smailagic and David Kogan. Location sensing and privacy in a context-aware computing environment. *IEEE Wireless Communications*, pages 10–17, 2002.
- [18] M. Brunato and C. K. Kalló. Transparent location fingerprinting for wireless services. In *Proceedings of Med-Hoc-Net 2002*, pages XXX–XXX, September 2002.
- [19] S.Y. Seidel and T.S. Rappaport. 914 mhz path loss prediction model for indoor wireless communications in multi-floored buildings. In *IEEE Transactions on Antennas & Propagation*, February 1992.
- [20] Katia Runser, Philippe Buhr, Guillaume de la Roche, and Jean-Marie Gorce. Validation de la méthode de prédiction de couverture radio mr-fdpf. In *Actes des 6e Rencontres Francophones AlgoTel*, pages 21 – 26, Mai 2004.
- [21] Moustafa Youssef, Ashok Agrawala, and Udaya Shankar. Wlan location determination via clustering and probability distributions. In *IEEE PerCom 2003*, March 2003.
- [22] Moustafa A. Youssef, Ashok Agrawala, A. Udaya Shankar, and Sam H. Noh. A probabilistic clustering-based indoor location determination system. Tech. Report CS-TR-4350, University of Maryland, March 2002.
- [23] M. Youssef and A. Agrawala. Handling samples correlation in the horus system. 2004.
- [24] Paul Castro, Patrick Chiu, Ted Kremenek, and Richard R. Muntz. A probabilistic room location service for wireless networked environments. In *UbiComp '01 : Proceedings of the 3rd international conference on Ubiquitous Computing*, pages 18–34, London, UK, 2001. Springer-Verlag.
- [25] Andreas Haeberlen, Eliot Flannery, Andrew M. Ladd, Algis Rudys, Dan S. Wallach, and Lydia E. Kavradi. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the Tenth ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Philadelphia, PA, September 2002.
- [26] Y. Wang, X. Jia, and H.K Lee. An indoors wireless positioning system based on wireless local area network infrastructure. In *6th Int. Symp. on Satellite Navigation Technology Including Mobile Positioning & Location Services*, number paper 54, Melbourne, July 2003. CD-ROM proc.
- [27] Inc Interlink Networks. A practical approach to identifying and tracking unauthorized 802.11 cards and access points. Technical report, 2002.

-
- [28] Jean Tourrilhes. Wireless tools. http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Tools.html, 1996.
 - [29] R. Battiti, A. Villani, and T. Le Nhat. Neural network models for intelligent networks : deriving the location from signal patterns, 2002.
 - [30] D. Charlet, F. Lassabe, P. Canalda, P. Chatonnay, and F. Spies. Mobility prediction for multimedia services. In I. K. Ibrahim and Johannes Kepler University Linz, editors, *Handbook of Research in Mobile Multimedia*, chapter 33, pages 491–506. Idea Group Inc., 2006. ISBN : 1591408660.
 - [31] F. Lassabe, O. Baala, Ph. Canalda, P. Chatonnay, and F. Spies. A friis-based calibrated model for WiFi terminals positioning. In *Proceedings of IEEE Int. Symp. on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005)*, pages 382–387, Taormina, Italy, June 2005.
 - [32] Arnaud Troël, Frédéric Weiss, and Michel Banâtre. Prise en compte du mouvement dans les systèmes de communication sans-fil. Technical Report 4720, INRIA, Février 2003.
 - [33] *Mobility Prediction Model for Cellular Networks Based on the Observed Traffic Patterns*, July 2002.
 - [34] S.J. Lee, W. Su, and M. Gerla. Mobility prediction in wireless networks. In *Proceedings of IEEE ICCCN*, pages 22–25, October 2000.
 - [35] Robert Grover Brown and Patrick Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, Inc., third edition, 1997.
 - [36] Joseph J. LaViola Jr. Double exponential smoothing : An alternative to kalman filter-based predictive tracking. In *Proceedings of the workshop on Virtual environments 2003*, pages 199–206. ACM Press, 2003.
 - [37] T. Liu, P. Bahl, and I. Chlamtac. Mobility modeling, location tracking and trajectory prediction in wireless atm networks. *IEEE Journal on Selected Areas in Communications*, 16(6) :922–936, August 1998.
 - [38] J. MONTAVONT, E. IVOV PETROV, and T. NOËL. Analysis of a geolocation-based fmipv6 extension for next generation wireless lans. *Ubiquitous Computing and Communication Journal*, 2(5) :10 pp, october 2007.
 - [39] Ramesh R. Sarukkai. Link prediction and path analysis using markov chains. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications networking*, pages 377–386, Amsterdam, The Netherlands, The Netherlands, 2000. North-Holland Publishing Co.
 - [40] Peter L. T. Pirolli and James E. Pitkow. Distributions of surfers’ paths through the world wide web : Empirical characterizations. *World Wide Web*, 2(1-2) :29–45, 1999.
 - [41] James Pitkow and Peter Pirolli. Mining longest repeating subsequences to predict world wide web surfing. In *USITS’99 : Proceedings of the 2nd conference on USENIX Symposium on Internet Technologies and Systems*, pages 13–13, Berkeley, CA, USA, 1999. USENIX Association.
 - [42] M. Deshpande and G. Karypis. Selective markov models for predicting web page accesses. *ACM Transactions on Internet Technology*, 4(2) :163–184, May 2004.

- [43] Mazen Tlais and Frédéric Weis. Mobility prediction in 4g d-cov networks. 2005.
- [44] J.M. François, G. Leduc, and S. Martin. Learning movement patterns in mobile networks : a generic method. In *European Wireless 2004*, pages 128–134, February 2004.
- [45] J.M. François, G. Leduc, and S. Martin. Evaluation d’une méthode de prédiction des déplacements de terminaux dans les réseaux mobiles. In Hermès Lavoisier, editor, *Réseaux mobiles et ad hoc, qualité de service, test et validation, ingénierie du trafic*, pages pp. 189–202. Colloque Francophone sur l’Ingénierie des protocoles, October 2003.
- [46] Jan Petzold, Andreas Pietzowski, Faruk Bagci, Wolfgang Trumler, and Theo Ungerer. Prediction of indoor movements using bayesian networks. In *Proceedings of the First International Workshop on Location- and Context-Awareness, LoCA 2005*, volume 3479, pages 211–222. Verlag : Springer Berlin / Heidelberg, May 2005.
- [47] Damien Charlet. *Mobilité et coopération dans un système réparti de caches vidéo*. PhD thesis, École doctorale SPIM, 2004.
- [48] D. Charlet, P. Canalda, P. Chatonnay, and F. Spies. Mobilité dans les caches vidéo : une stratégie adaptative. première soumission à Technique et science informatiques, le 20 Avril 2004, 2004.
- [49] Julien Bourgeois, Emmanuel Mory, and François Spies. Video transmission adaptation on mobile devices. *Journal of Systems Architecture*, 49 :475–484, 2003.
- [50] J. Petzold. Augsburg indoor location tracking benchmarks. Tech. Report 2004-9, Institute of Computer Science, University of Augsburg, Germany, February 2004.
- [51] J. Bourgeois, E. Mory, and F. Spies. Video transmission adaptation on mobile devices. *Journal of Systems Architecture*, 49(10-11) :475–484, November 2003.
- [52] B. Mac Larnon. *TAPR’s Spread Spectrum Update*, chapter VHF/UHF/Microwave Radio Propagation : A Primer for Digital Experimenters. Tucson Amateur Packet Radio Corporation, 1998.
- [53] Paul Mühlethaler. *802.11 et les réseaux sans fil*. Eyrolles, August 2002.
- [54] Dominique Dhoutaut. *Etude du standard 802.11 dans le cadre des réseaux ad-hoc : de la simulation à l’expérimentation*. PhD thesis, Ecole Nationale Supérieure de Lyon, December 2003.
- [55] P. Karn. Maca - a new channel access method for packet radio. In *ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, pages 134–140, 1990.
- [56] William Stallings. *Réseaux et communication sans fil*. Pearson Education, 2 edition, 2005.
- [57] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 1991.
- [58] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. volume 77, pages pp. 257–286. IEEE, February 1989.
- [59] David M. Bourg and Glenn Seeman. *AI for Game Developers*. O’Reilly, July 2004.
- [60] J. Borges and M. Levene. An average linear time algorithm for web usage mining. *International Journal of Information Technology & Decision Making*, 3(2) :307–319, June 2004.
- [61] Jose Borges and Mark Levene. Data mining of user navigation patterns. In *WEBKDD*, pages 92–111, 1999.

-
- [62] José Borges and Mark Levene. A dynamic clustering-based markov model for web usage mining. *Computer Research Repository*, cs.IR/0406032, May 2004.
 - [63] Thanh-Hà Lê. Localisation des stations mobiles dans les réseaux sans-fil. Master's thesis, INSA Lyon, September 2004.
 - [64] G. D. Forney. The viterbi algorithm. In *Proceedings of the IEEE*, volume 61, pages 268–278, March 1973.
 - [65] J. Tang, G. Xue, and W. Zhang. Reliable routing in mobile ad hoc networks based on mobility prediction. In *Mobile Ad Hoc and Sensor Systems*, 2004.
 - [66] J.M. François and G. Leduc. Mobility prediction's influence on qos in wireless networks : a study on a call admission algorithm. In *Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'05)*, pages 238–247, 2005.
 - [67] Arnaud Troël, Frédéric Weis, and Michel Banâtre. Prise en compte du mouvement dans les systèmes de communication sans-fil. Rapport de recherche 4720, INRIA Rennes, February 2003.
 - [68] S.O. Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, 23 and 24, 1944 and 1945.
 - [69] H. Hashemi. The indoor radio propagation channel. In *Proceedings of the IEEE*, volume 81, pages 943–968. IEEE, July 1993.
 - [70] R.J. Bates. *GPRS : General Packet Radio System*. McGraw-Hill, New York, 2002.
 - [71] Andrew Tanenbaum. *Réseaux*. InterEditions, 1997.
 - [72] Vincent Huberty. La géolocalisation des gsm. [http ://www.astel.be/article585.html](http://www.astel.be/article585.html).
 - [73] Q.931 protocol overview. [http ://www.freesoft.org/CIE/Topics/126.htm](http://www.freesoft.org/CIE/Topics/126.htm).
 - [74] Mohamed-Slim Alouini. Global positioning system : An overview. *Tunisian Scientific Magazine*, 10(1) :49–51, december 1996.
 - [75] Les soldats du futur. [http ://www.gign.org/dossiers/mai2003.php](http://www.gign.org/dossiers/mai2003.php), 2003.
 - [76] R.B. Marth, R. Levi, C.T. Judd, I.N. Duboraw, and K. Beam. The integrated navigation capability for the force xxi land warrior. Technical Report A269453, ARMY TOPOGRAPHIC ENGINEERING CENTER FORT BELVOIR VA, 1998.
 - [77] University of Paderborn. Mupad research group. [http ://www.mupad.de](http://www.mupad.de), January 2004.
 - [78] Guy Pujolle. *Les réseaux*. Eyrolles, 2003.
 - [79] S. Lee, W. Su, and M. Gerla. Ad hoc wireless multicast with mobility prediction, 1999.
 - [80] J. Julliand. Automates finis et applications. Cours de DEA IAP, 2003.
 - [81] Bruno Baynat. *Théorie des files d'attente*. Hermès Science Publications, 2000.

Résumé

La démocratisation des terminaux mobiles et l'accroissement des débits disponibles permettent d'envisager de nouvelles applications, en particulier relatives au contexte. Celles-ci nécessitent d'assurer la continuité des services et la détection de la position du terminal mobile. Nous proposons d'une part la géolocalisation des terminaux et, d'autre part, la prédiction de la mobilité.

Les systèmes satellites ne fonctionnant pas à l'intérieur des bâtiments, nous basons nos travaux sur les réseaux Wi-Fi. Deux méthodologies se démarquent pour localiser un terminal Wi-Fi : l'une repose sur une cartographie des puissances, l'autre repose sur le calcul des distances entre le terminal et des points dont les coordonnées sont connues. Chaque modèle ayant ses points faibles, nous les avons combinés pour améliorer la précision finale.

Nous proposons un premier modèle qui calcule les distances entre le terminal mobile et des points d'accès en se basant sur la puissance du signal reçu. Il en déduit la position du terminal par calcul. Le second modèle proposé restreint la recherche à une zone homogène grâce à la cartographie des puissances avant d'utiliser le premier modèle.

Nous avons expérimenté nos modèles ainsi que les modèles fondamentaux de l'état de l'art en étendant leurs conditions d'application. Les résultats des systèmes basés sur la propagation des ondes sont de l'ordre de 9 à 15 mètres d'erreur. Les modèles basés sur une cartographie permettent quant-à-eux d'atteindre une précision de l'ordre de 3 à 7 mètres selon les conditions.

L'historique des positions permet à un système d'apprentissage d'acquérir un modèle des déplacements des terminaux puis de prédire les déplacements futurs par l'étude et la comparaison du modèle obtenu à des déplacements ultérieurs. Nous avons proposé en particulier d'employer les chaînes de Markov et les réseaux bayésiens pour effectuer l'apprentissage et la prédiction de la mobilité. Nous avons enrichi ces modèles d'un seuil qui détermine le choix des politiques à appliquer en fonction des déplacements du terminal. La précision de nos modèles est variable en fonction des paramètres d'ordre et de seuil mais permet d'atteindre des taux de réussite de la prédiction de 75%. Cette précision permet d'envisager l'anticipation des handovers et l'application d'une politique appropriée.

Mots-clés: mobilité, continuité de service, système pervasif, prédiction, géolocalisation

Abstract

Democratization of mobile terminals and bandwidth capability allow to consider new applications, particularly context-aware applications. Such applications require service continuity and mobile terminal positioning. We propose on one hand to locate mobile terminals and, on the other hand, mobility prediction.

Navigation Satellites Systems are not working indoors. Thus we base our work on IEEE 802.11 networks. Two methods exist to locate a Wi-Fi terminal : the first one is based on a signal strength map. The other one is based on computation of distances between the mobile terminal and points whose coordinates are known. Each method having its own drawbacks, we merge both of them to improve positioning accuracy.

We propose a first model, which computes distances between the terminal and the access points, based on the SS received. Terminal's location is inferred by calculation. The second model we propose restricts the positioning to an area through an SS map. Then, in this area, the first model is applied to determine the terminal's location.

We tested our models and some models we studied, varying the tests conditions. Distance computation-based systems achieve an accuracy from 9 to 15 meters. The SS map-based ones reach an accuracy from 3 to 7 meters.

Locations history allows a learning system to build a mobile terminals mobility model, allowing to predict further moves by comparing new moves to the model. We propose to model mobility through Markov models and bayesian networks. We add a threshold to these models to determine a mobility-related policy to the terminal. Accuracy of the models vary according to the threshold value and the order of the Markov model. However, the models reach 75% good guesses when trying to predict a terminal's move. Such accuracy allows to consider handover anticipation by applying an adequate policy.

Keywords: mobility, service continuity, pervasive computing, prediction, positioning